# Convective Initiation – Relevant Processes and Their Representation in Convection-Permitting Models

**Mirjam Johanna Hirt**


München - 2020

# Convective Initiation – Relevant Processes and Their Representation in Convection-Permitting Models

**Mirjam Johanna Hirt**

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Mirjam Johanna Hirt
aus Friedberg

München, 7. April 2020

Mirjam Hirt: *Convective Initiation – Relevant Processes and Their Representation in Convection-Permitting Models,* PhD thesis in Physics, April 2020

ERSTGUTACHTER: Prof. Dr. George C. Craig
ZWEITGUTACHTER: Prof. Dr. Mark Wenig
DATUM DER ABGABE: 07.04.2020
DATUM DER MÜNDLICHEN PRÜFUNG: 24.06.2020

Parts of this thesis are contained in:

**Hirt, M.**, Rasp, S., Blahak, U., & Craig, G. (2019). Stochastic parameterization of processes leading to convective initiation in kilometre-scale models. *Monthly Weather Review*, 147, 3917–3934, doi: 10.1175/MWR-D-19-0060.1. ©American Meteorological Society. Used with permission.

**Hirt, M.**, Craig, G. C., Schäfer, S., Savre, J., & Heinze, R. (2020). Cold pool driven convective initiation: using causal graph analysis to determine what km-scale models are missing. *Quarterly Journal of the Royal Meteorological Society*, doi: 10.1002/qj.3788.

# ZUSAMMENFASSUNG

Die Vorhersage von konvektivem Niederschlag ist für heutige, numerische Wettervorhersagemodelle eine besondere Herausforderung. Die Ursache dafür liegt insbesondere an unzureichend aufgelösten Grenzschichtprozessen, welche für die Entstehung von Konvektion besonders relevant sind. Diese Arbeit befasst sich mit drei solchen Prozessen, die Grenzschichtturbulenz, mechanische Hebung durch Subgrid-skalige Orographie und Cold Pools. Ziel ist es, systematische Fehler in den simulierten Prozessen zu identifizieren und die Darstellung der Prozesse in Konvektions-erlaubenden Modellen zu verbessern.

Grenzschichtturbulenz wird in den meisten Modellen durch Parameterisierungen näherungsweise dargestellt, indem nur der mittlere Einfluss der Turublenz in einer Gitterbox berücksichtigt wird. Für die Entstehung von Konvektion ist jedoch die subgrid-skalige Variabilität von entscheidender Bedeutung. Um diese fehlende Variabilität dennoch zu berücksichtigen, haben Kober und Craig (2016) ein physikalisch basiertes, stochastisches Störungschema (PSP) entwickelt, wodurch zusätzlich Modellunsicherheiten direkt dort quantifiziert werden, wo sie entstehen. Das ist besonders wichtig, um zuverlässige Vorhersagen zu erhalten. Das Verhalten des PSP-Schemas zeigt jedoch in manchen Fällen unerwünschte Effekte, welche wir in dieser Arbeit verringern. Dafür entwickeln wir einige Modifikationen des PSP Schemas. Zum Beispiel stören wir die horizontalen Wind-Komponenten, sodass 3d Divergenzfreiheit gegeben ist und die Vertikalwindstörungen länger anhalten. Daraus entsteht eine entsprechend verbesserte Version, das PSP2, welches stärker unserem physikalischen Verständnis folgt und ähnliche Verbesserungen beim Einsetzen von konvektivem Niederschlag aufweist wie in dem ursprünglichem PSP-Schema und zusätzlich eine verbesserte Strukture der Niederschlagszellen zeigt.

Als nächst-wichtigsten Prozess für die Konvektionsauslösung berücksichtigen wir die Hebung durch subgrid-skalige Orographie. Dafür entwickeln wir eine weitere, stochastische Parameterisierung, das SSOSP-Schema, welches den Effekt der mechanischen Hebung durch subgrid-skalige Orographie darstellt. Dazu wird die Amplitude von Schwerewellen durch Informationen zur subgrid-skaligen Orographie dargestellt und verwendet. Zwar ist durch das SSOSP-Schema eine deutliche Erhöhung der Konvektionsauslösung in orografischen Regionen möglich, jedoch fällt dies mit einem unerwünschten Einfluss auch auf nicht-orografische Regionen zusammen. Im Gegensatz zu unseren ursprünglichen Erwartungen kommen wir zu dem Schluss, dass die subgrid-skalige Orographie keine entscheidende Rolle spielt, da sie meist von aufgelöster Orographie begleitet wird.

Der bedeutendste Teil der Arbeit befasst sich mit Cold Pools. Cold Pools sind vor Allem für die Organisation von Konvektion entscheidend, sowie für die Auslösung von Konvektion am Nachmittag und Abend. Dass Konvektions-erlaubende Modelle Cold Pools nicht ausreichend gut darstellen können, liegt nahe. Um genauer zu verstehen, welche Aspekte der Cold Pools mangelhaft dargestellt werden, verwenden wir hochauflösende Simulationen um Cold Pools, Cold Pool Ränder und ausgelöste Konvektion zu identifizieren. Dabei stellen wir fest, dass Cold Pools in Simulationen mit niedrigerer Auflösung häufiger, kleiner und weniger intensiv sind und schwächere Böenfronten aufweisen. Wir verwenden eine lineare Kausalitätsanalyse, um verschiedene indirekte Effekte zu quantifizieren. Dabei finden wir einen dominierenden Effekt: Durch die Reduzierung der Gittergrößen wird der aufwärts-gerichtete Massenfluss an der Böenfront reduziert, was zu geringeren Wahrscheinlichkeiten für die Konvektionsauslösung führt.

Basierend auf diesem gewonnen Verständnis entwickeln wir anschließend eine deterministische Cold-Pool-Parametrisierung, CPP, die die Aufwärtsbewegung an den Böenfronten von Cold-Pools verstärkt, um die Konvektionsauslösung durch Cold Pools zu verbessern. Dabei wird der Niederschlag besonders am Nachmittag und Abend verstärkt und die Organisation der Konvektion nimmt zu, woruch Niederschlagsvorhersagen verbessert werden.

# ABSTRACT

Current deficits of numerical weather prediction models in predicting convective precipitation are presumably caused by insufficiently resolved boundary-layer processes and their role in initiating convection. This thesis addresses three such boundary-layer processes, which are expected to be the most relevant ones for convective initiation. We identify current deficits and improve their representation in convection-permitting models to improve the representation of convection itself. The three considered processes are boundary-layer turbulence, mechanical lifting by subgrid-scale orography and cold pools.

Boundary layer turbulence is mostly parameterized in these models and only represented by the average effect on a grid box. As the subgrid-scale variability is crucial for convective initiation, Kober and Craig (2016) have developed a unique, physically based stochastic perturbation scheme (PSP) to reintroduce the subgrid-scale variability of boundary-layer turbulence in a stochastic manner. This allows the quantification of model uncertainty at its source - a crucial, but rarely performed step for reliable forecasts. As the PSP scheme also introduced some undesired behavior, we develop several modifications to pertain the positive effect of the original PSP scheme while being physically more consistent. For example, we included perturbations in horizontal wind components in a 3d non-divergent way to obtain persistent vertical velocity perturbations. A revised version, PSP2, is physically more consistent and shows improvements in the onset of convective precipitation similar to the original scheme and an improved structure of precipitation cells.

The next important process is assumed to be convective initiation by subgrid-scale orography. We develop an additional stochastic parameterization, the SSOSP scheme, to include the effect of mechanical lifting by subgrid-scale orography using a gravity wave formalism and information on subgrid-scale orography. While a clear increase in convective initiation over orographic regions is possible by the SSOSP scheme, this usually coincides with an undesired impact also on non-orographic regions. We conclude - in contrast to our initial expectations - that, most likely, subgrid-scale orography does not play a crucial role because it is mostly accompanied by resolved orography.

The most notable parts of this thesis are concerned with cold pools. Cold pools are expected to play a crucial role in the convective organization and for late afternoon and evening convective initiation. However, km-scale models are not expected to simulate cold pools with sufficient accuracy. To better understand, which aspects of cold pools are insufficiently represented, we identify cold pools, cold pool boundaries and initiated convection in high-resolution simulations. We find that cold pools are more frequent, smaller, less intense and have weaker gust fronts in lower resolution simulations. We use a linear causal graph analysis to disentangle different indirect effects. Doing so, we identify one single, dominant effect: reducing grid sizes reduces upward mass flux at the gust front directly, which causes weaker triggering probabilities.

Based on these results, we then develop a deterministic cold pool parameterization, CPP, which strengthens the upward motion at cold pool gust fronts to improve cold pool driven convective initiation. We find that precipitation is amplified and becomes more organized in the afternoon and evening. Better precipitation forecasts will then be possible.

# CONTENTS

# 1 INTRODUCTION

Atmospheric deep convection, or simply convection, represents the formation of fast-developing cloud systems, that extend throughout the whole depth of the troposphere. The most prominent example of such clouds are thunderstorms. Such convective clouds can result in high wind speeds, heavy precipitation, hail or flooding, thereby disturbing our everyday life, devastating areas and causing material losses and even casualties. In June 2019, for instance, a specifically large and long-lived thunderstorm occured over Bavaria which caused hail with up to 5 cm diameter, extreme precipitation ($27\,mm\,h^{-1}$), and wind speeds of up to $120\,km\,h^{-1}$ (Deutscher Wetterdienst, 2019). According to Munich Re (2020), damages of almost 1 billion dollars occured, which made this single thunderstorm event the most expensive natural hazard in Germany in 2019. A different, highly convective weather situation caused severe flooding and damages in several towns in Germany in May and June 2016. These floodings lead to financial losses of up to 2.2 billion dollars (Piper *et al.*, 2016; Kron *et al.*, 2019). To prevent, or at least mitigate, such tremendous damages, people, companies or governments require reliable warnings and, hence, predictions of such convective events sufficiently in advance.

However, predicting convective clouds has been a long-standing challenge. Until the last century, the prediction of thunderstorms was based on eye-observations of clouds, the categorization of previous weather types and experience (Nebeker, 1997). Consequently, only very short-term or relatively crude predictions were possible. When the first numerical weather predictions became operational in the 1950s and 60s (Golding *et al.*, 2004; Harper *et al.*, 2007; Randall *et al.*, 2018), it was finally possible to predict the larger-scale weather based on physical equations, which tremendously improved weather forecasts. However, the lack of computational power, model complexity and observations still prohibited the explicit simulation of deep convective clouds for several decades. In the late 1980s, with the availability of radar and satellite observations, nowcasting techniques were developed to project the behavior of existing clouds into the future (Golding *et al.*, 2004). These extrapolation techniques, however, could not predict the initiation of new cells, nor were their forecasts skillful beyond a few hours.

A milestone was reached when increasing computational possibilities joined with intensive model development enabled the explicit simulation of deep convective circulations with numerical models in the 1990s. At first, such simulations were only feasible for research purposes and selected events (Clark *et al.*, 2016). Lilly (1990) speculated about the operational, numerical prediction of thunderstorms in 1990, but it still took more than a decade for most weather services to develop and afford models suited for this challenge (e.g., Baldauf *et al.*, 2011; Clark *et al.*, 2016). Today, such convection-permitting models are the model type of choice for regional weather prediction and many other applications.

While these models have undoubtedly led to a step-change in forecasting deep convection (Clark *et al.*, 2016), several deficits remain. The primary deficits in simulating deep convective systems manifest in biases in the diurnal cycle and a lack of organization of convection. Responsible are processes, that are tightly coupled to convection, but are so small, that even convection-permitting models cannot resolve them explicitly. This grey zone for convection modeling poses a severe challenge for current weather prediction. The largest portion of these deficits is ascribed to insufficiently represented processes in the boundary layer ($\approx$ the lowest kilometer of the atmosphere), which are crucial for initiating the formation of deep convective clouds. The presumably three most essential processes for improving convective initiation are boundary-layer turbulence, small-scale orography and cold pools. In this thesis, we will better understand which aspects of these processes are insufficiently represented in numerical models and develop new or recently proposed approaches to improve their representation. Doing so, we address the current challenge of the convective grey zone.

## 1.1 NUMERICAL WEATHER PREDICTION

Numerical weather prediction (NWP) uses computational models that describe atmospheric flow to a certain degree of complexity. They consist of physical, partial differential equations, including the momentum equations for each velocity component, the thermodynamic equation, the continuity equation and prognostic equations for moisture quantities. As they cannot be solved analytically, these equations are approximated for a discrete three-dimensional grid representing the earth's atmosphere. First, a current state of the atmosphere, i.e., the initial condition, has to be derived from observations and previous forecasts, a process termed data assimilation. Then numerical integration for future, discrete time steps enables a prediction of the future state of the atmosphere. Processes that occur on scales smaller than the model grid spacing, i.e., subgrid-scale processes, are therefore not explicitly included. Instead, their impact is approximated in terms of the resolved variables and parameters. Such approximations are called parameterizations and will be introduced in more detail in Sec. 1.2 and 1.3.

### 1.1.1 *Errors and uncertainty*

The fundamental challenge of NWP originates due to the atmosphere's chaotic nature - a characteristic made popular by the butterfly effect - which is intimately connected with a myriad of atmospheric scales interacting with each other. As a consequence, small errors, e.g., in the initial conditions, grow rapidly in space and time and are hence projected on all scales. This error growth radically limits our ability to predict the weather to a few weeks, days and sometimes even hours[1].

1 Often, a clear distinction between intrinsic and practical predictability is made. The former describes for how long in advance useful, hypothetical predictions can be made given infinitesimal errors. Despite the deterministic nature of the governing equations, this time scale does not grow indefinitely with decreasing errors. The practical predictability, on the other hand, is a characteristic of current NWP systems with finite errors. The specific limit of both types of predictability strongly depends on the spatial scales of interest and the flow situation.

Errors originate from all parts of the NWP chain, for instance, from insufficient observations, from erroneous initial conditions and approximations in the numerical methods and physical parameterizations. Progress in reducing such errors can be made by increasing computational power, improving different components of NWP models, or by intensive testing and tuning. This last aspect, although perhaps unsatisfying from a scientific point of view, is commonly done at operational weather services and refers to the determination of otherwise vague parameters by optimizing forecast skill scores (Mauritsen *et al.*, 2012; Hourdin *et al.*, 2017).[2]

As it is not possible to completely remove errors, and as substantial biases still exist in real-world models, their impact on the forecast should be quantified. This uncertainty quantification is now typically done using ensembles (Buizza, 2018) by perturbing each ensemble member within the range of uncertainty. Different methods to generate ensemble members include perturbations in initial and boundary conditions, parameter perturbations of parameterizations, multi-physics approaches where different parameterization schemes are used, or even the use of different models. In addition, stochastic parameterizations (see Sec. 1.2.2) inherently include uncertainty and are also often used to generate ensembles. By using such ensemble methods, uncertainty is quantified, and its nonlinear, flow-dependent propagation in space and time is simulated.[3] Overall, reducing and quantifying uncertainty is continuously being achieved at various components of NWP. As a result, the practical predictability has increased by one day per decade since the 1980s, which even gained the notion of a "quiet revolution" (Bauer *et al.*, 2015, title).

### 1.1.2 *Challenges for convective scale NWP*

Simulating and predicting convective processes, and thus convective precipitation, is a particular challenge for NWP. The following four aspects dominate this challenge. First, convection is "fundamentally a turbulent process" (Yano *et al.*, 2018, p. 701), and is substantially more nonlinear than synoptic-scale systems. Hohenegger and Schär (2007), for instance, compare a 7 h convection forecast to a 10-day synoptic forecast concerning nonlinearity. Consequently, errors grow more rapidly, and predictability of convective scale systems is limited to only a few hours (Hohenegger and Schär, 2007; Trapp, 2013). Additionally, the assimilation of convective scale observations poses a current challenge (see Gustafsson *et al.*, 2018, for a review). Furthermore, due to the intermittent and often localized behavior of convection, traditional verification metrics often fail to give meaningful evaluations, which hinders model development. Alternative approaches include fuzzy or object-based verification methods. Some will be introduced

2 Although the necessity to tune the numerical models is widely acknowledged, tuning can result in different biases compensating each other and models might yield the right results for the wrong reasons. Such compensating errors can substantially hinder model development as an improved representation of physical processes can deteriorate forecast skills when biases are not compensated any more (Palmer and Weisheimer, 2011). Furthermore, models are sometimes tuned to optimize single variables at the expense of others (Baldauf *et al.*, 2011). We will come back to tuning related problems throughout the thesis.

3 An alternative approach to quantify and predict uncertainty is based on the prediction of the Liouville equation, as the fundamental quantity for uncertainty. However, its application in operational, fully complex models is not feasible (Ehrendorfer, 1994a,b; Buizza, 2018).

in Chapter 2. Another major complication arises as many scales relevant for convection are so small that they are often only poorly represented by the model dynamics or the parameterizations. These include evaporation, droplet formation, entrainment, the initiation of convection and sometimes even the deep convective circulations themselves. This aspect will become more evident in the following section.

### 1.1.3 *Atmospheric models of different scales*

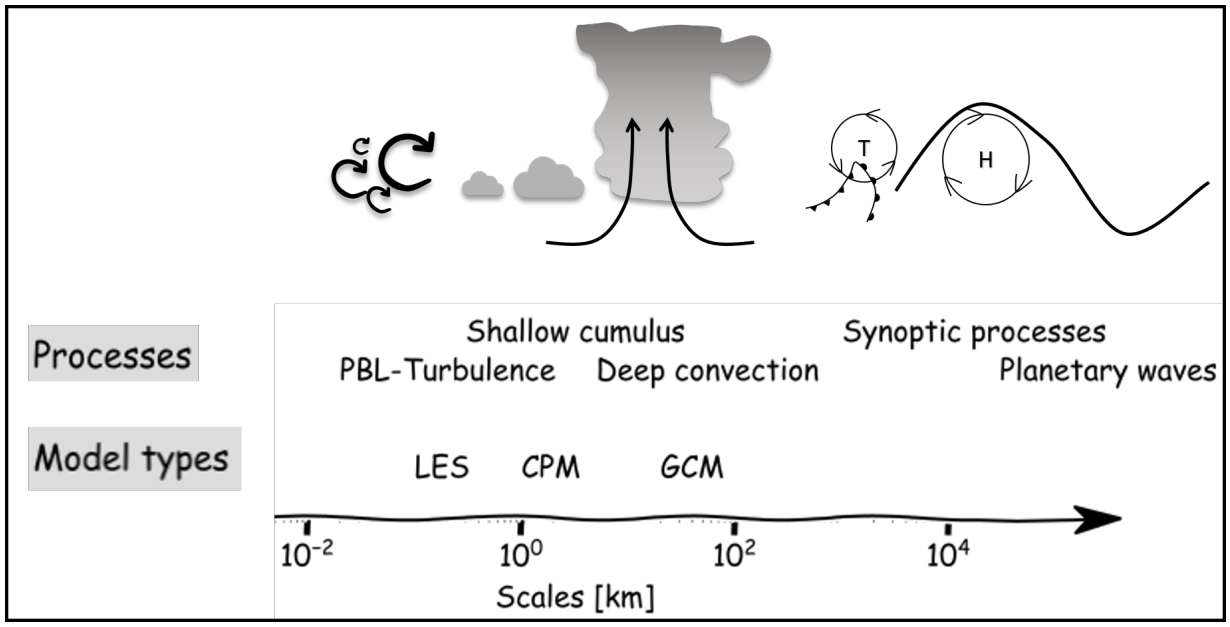


**Figure 1:** Illustration of processes at different scales in comparison to different model types and their grid sizes. As described in the text, to explicitly resolve the processes, the grid sizes of the models need to be significantly smaller than the process itself. Consequently, LES can resolve larger turbulent eddies and shallow convection, convection-permitting models (CPMs) can resolve larger deep convective circulations and general circulation models (GCMs) only synoptic or planetary processes.

The atmosphere contains many different processes of different spatial and temporal scales. Some characteristic phenomena of different scales are turbulence, thunderstorms/deep convection or synoptic high and low-pressure systems, to name a few. Turbulence in the boundary layer has its largest spatial scales in the range of the boundary-layer height (usually less than 1 km) and a typical lifetime of several minutes. Deep convective circulations, on the other hand, extend few kilometers and few hours. Synoptic high or low-pressure systems have even larger spatial scales of thousand kilometers and usually last for a few days.

Depending on the processes of interest, different types of numerical models are being used. For numerical models to explicitly simulate any process, their grid sizes have to be significantly smaller than the process itself. Otherwise, the process can only be approximated by parameterizations (see Sec. 1.2), resulting in a crude approximation of the effect on grid scales. But computational resources are limited, and a mere halving of grid size easily octuples the computational requirements (e.g., Warner, 2010,

p. 362). Therefore, the computational requirements of smaller grid sizes are compensated by reducing computational requirements otherwise. Usually, this is obtained by reducing total domain size, but limiting total integration time or ensemble dimension reduces computing requirements as well. This tradeoff gives rise to various types of atmospheric, numerical models depending on their purpose and the respective processes that they can simulate explicitly.

We illustrate this hierarchy of atmospheric models by considering three types of models as selected examples, namely global NWP models or general circulation models (GCMs), convection-permitting models (CPMs) and large eddy models for turbulent scales (LES). See also Fig. 1 for the grid sizes of these models and corresponding processes.

Global weather prediction models, which simulate the global atmosphere of the earth, usually can only afford grid sizes of ten kilometers. For instance, the operational setup from the German weather service (DWD), has an effective grid spacing of 13 km (Reinert *et al.*, 2018) and the European center for medium-range weather forecasting (ECMWF) uses a discretization of approximately 9 km and 18 km for their high-resolution run and their ensemble, respectively (ECMWF, 2019). Such resolutions are sufficient to simulate synoptic fronts, cyclones or other high- and low-pressure systems or planetary waves explicitly.

To resolve thunderstorm circulations or larger convective clouds, grid sizes of less than a few kilometers are necessary, often 1-5 km grid sizes are used. We will refer to such models as km-scale or *convection-permitting models* (CPM). This regime of models is now widely used for different applications: convection-permitting models are used for regional weather prediction by most weather services (Baldauf *et al.*, 2011; Clark *et al.*, 2016); they are being tested for applications in regional climate modeling (Leutwyler *et al.*, 2017; Schär *et al.*, 2019); and they are even used for exploratory, global convection-permitting simulations (Judt, 2018; Stevens *et al.*, 2019b; Zhou *et al.*, 2019). Convection-permitting models will doubtlessly be used for even more applications in the future. For instance, at ECMWF, global models are planned to have 5 km grid sizes by 2025 (ECMWF, 2016), and Palmer (2019a) is speculating that global ensemble simulations will have 1 km grid sizes in 25 years.

On even smaller scales, large-eddy simulations (LES) can be used to simulate larger turbulent eddies in the planetary boundary layer explicitly. These models require grid sizes well below 1 km, and sometimes have grid sizes of only tens of meters. LES are often used for idealized cases, but also more realistic situations have been computed for research purposes (Heinze *et al.*, 2017; Stevens *et al.*, 2019a). As far as we are aware, no plans exist to make such models operational in the next decade.

Hence, models with grid sizes of few kilometers will likely dominate the atmospheric modeling landscape for years to come. Maximizing the highly flow-, space- and time-dependent predictability of convection within such km-scale models is consequently a dominant challenge for NWP today, and much research is devoted towards it. This includes coupling subgrid-scale processes to the resolved convective scales and quantifying uncertainties accordingly. This will be a challenge throughout this thesis, where we specifically focus on the improved representation of convective initiation by subgrid-scale or insufficiently resolved processes. Such coupling of resolved and unresolved

processes can be achieved by parameterizations, which we will formally introduce in the following.

## 1.2 PARAMETERIZATIONS

Parameterizations approximate the effect of subgrid-scale, unresolved processes on the resolved scale. Let us consider the set of variables $\phi$ averaged over the model grid box, $\overline{\phi}$. The residual, i.e., the subgrid-scale fluctuations from the grid mean, is represented as $\phi'$, so that $\phi = \overline{\phi} + \phi'$. This partitioning is often denoted as Reynolds averaging.[4] Then numerical weather prediction can be conceptualized as the numerical integration of the prognostic equation for $\overline{\phi}$:

$$\frac{\partial \overline{\phi}}{\partial t} = f(\overline{\phi}) + \overline{g(\overline{\phi}, \phi')}, \tag{1}$$

where $f(\overline{\phi})$ corresponds to resolved scale dynamics and $\overline{g(\overline{\phi}, \phi')}$ represents interactions between resolved and subgrid-scales and how they impact the resolved scale variables. However, the subgrid-scale fluctuations $\phi'$ are not specified and $\overline{g(\overline{\phi}, \phi')}$ needs to be approximated only in terms of $\overline{\phi}$, i.e., $\overline{g(\overline{\phi}, \phi')} \approx h(\overline{\phi})$. This approximation is called parameterization.

A wide variety of processes is usually parameterized separately. The impact of unresolved eddies is approximated in turbulence parameterizations. The influence of shallow and deep convection on resolved scales is usually considered separately in parameterizations. Cloud processes, like condensation or droplet formation, are represented by microphysics parameterizations. Radiative processes are approximated for computational efficiency in radiation parameterization. Furthermore, many other processes like gravity wave drag, land surface processes or atmosphere-ocean interactions are often represented in some form of parameterization. In addition, interactions between different parameterizations also need to be included.

Various approaches are used to build parameterizations. Traditional parameterizations are usually based on empirical methods and conceptual process understanding. Sometimes parameterizations can be derived with several approximations from first principles, although empirical closure assumptions usually have to be made. Consequently, such parameterizations represent severe sources of errors and uncertainty in weather and climate models, and considerable effort is made to improve their accuracy. Progress with traditional parameterizations is ongoing, but also different perspectives are being considered (Yano, 2016; Rio *et al.*, 2019).

In the last years, the development of other approaches has grown. Exceptionally accurate are *Super-* or even *Ultraparameterizations*, which embed higher resolution models within each grid box of the coarser model (Grabowski, 2004; Parishani *et al.*, 2017; Ran-

4 Reynolds averaging represents the process of averaging equations after separating each variable $x$ into a slowly varying mean $\overline{x}$ and a fluctuating perturbation term $x'$: $x = \overline{x} + x'$. The average can be considered either over time, space or an ensemble (Wyngaard, 2010). See e.g., Markowski and Richardson (2011) for the fundamental three averaging rules and application.

dall *et al.*, 2018)[5]. Per large scale time step, the subgrid-scale tendencies can then explicitly be computed, which enables a high precision of these tendencies. The major drawback is the immense computational power required for the embedded high-resolution simulations. Hence, their application in operational modeling is not useful. Approaches based on deep neural networks are now also being developed. The neural networks are trained with traditional parameterizations, CRMs or Superparameterizations to emulate their behavior while being computationally more efficient (Gentine *et al.*, 2018; Rasp *et al.*, 2018a; Brenowitz and Bretherton, 2018). Several challenges, however, remain to be solved. Recently, so-called multi-fluid representations have been tested in meteorological applications, where the subgrid fluid is conditionally decomposed into multiple components (e.g., updrafts or downdrafts) (Thuburn *et al.*, 2018, 2019). More and more popular are stochastic parameterizations, where stochastic components are included in parameterizations to quantify uncertainty. They will play a major role in this thesis, and further details will be given in Section 1.2.2.

### 1.2.1 *Scale separation and grey zones*

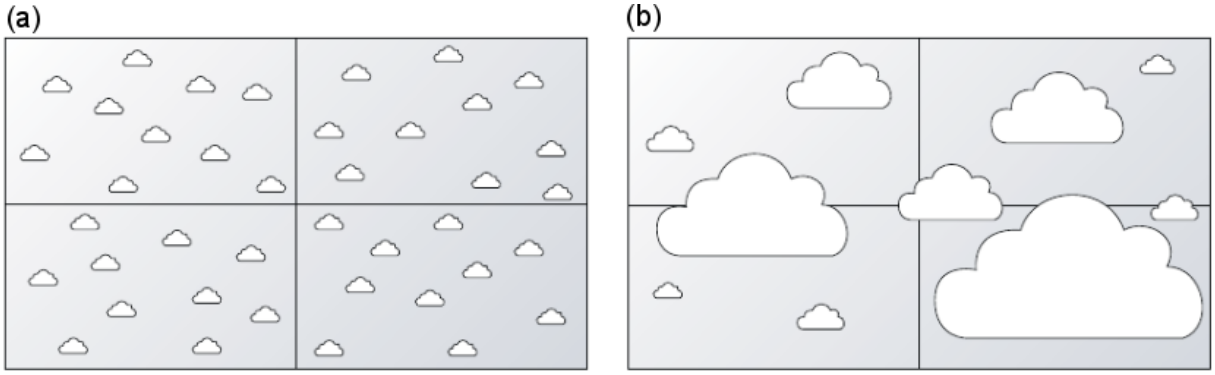


**Figure 2:** Illustration of the scale separation of convection parameterizations. The boxes represent model grid boxes. (a) displays the situation where the cloud scales and grid scales are well separated. In (b), the grey zone situation is displayed where some clouds are almost as large as the grid boxes themselves. From Palmer (2019a).

A fundamental problem for parameterizations is the artificial separation between the resolved and parameterized scales. For instance, in Fig. 2b, we can assume that the large-scale, synoptic conditions are identical for all four grid boxes, but the clouds are not. We see that there is a distribution of different cloud sizes where individual clouds can be as large as the grid size, while other clouds are too small to be explicitly resolved. Furthermore, some clouds can be attributed to more than just one grid box. Also, the lifetimes of these clouds may or may not be longer than the model time step. Despite identical large scale conditions, the average effect of the clouds on each grid box is not necessarily identical. Instead the impact on each grid box can be viewed as one realization of a distribution for $\overline{g(\overline{\phi}, \phi')}$ (see Equ. (1)), as conceptually illustrated in

5 Superparameterizations usually embed convection-permitting simulations with grid sizes of few kilometers within in the coarser model. Recently, LES-type models with grid sizes of only a few hundred meters have been embedded in a coarser model as Ultraparameterization (Parishani *et al.*, 2017).

Fig. 3 (green or orange curve). The figure further shows that the subgrid source terms are highly sensitive to the total number of clouds per grid box and hence strongly depend on the grid size ($N_{grey-\alpha}$ and $N_{grey-\beta}$). Craig and Cohen (2006) introduced a formal theory for such probability density functions for convective mass fluxes based on statistical mechanics.

These observations can be generalized for other processes as well, including turbulent eddies in the planetary boundary layer. An ideal parameterization should represent the probability distribution of underlying processes, automatically adapt to changing grid sizes, and should be correlated in space and time according to the spatial and temporal characteristics of the considered processes. These are, however, serious requirements for parameterizations and pose a severe challenge for numerical modeling today.

Traditional parameterizations were originally developed for models with tens or hundreds of kilometers. By assuming that the resolved scale acts on asymptotically larger spatial and temporal scales than the parameterized processes, traditional parameterizations circumvent the above described problems. Then, a scale-separation is obtained, as illustrated in Figure 2a for clouds. The subgrid-scale processes are temporally and spatially uncorrelated at the model temporal and spatial discretization, and no prognostic subgrid-scale variables, temporal or spatial correlations are required. Furthermore, following the central limit theorem, the distribution of the subgrid term $\overline{g(\bar{\phi}, \phi')}$ converges towards a Dirac delta function in the limit of infinitely many subgrid-scale objects as conceptually illustrated in Fig. 3. Consequently, the ensemble average approximates the subgrid-scale effects sufficiently well, which justifies traditional, deterministic parameterizations.

With the rise of convection-permitting models (see Sec. 1.1.3), these assumptions are now strongly violated for convection or turbulence, while the respective processes are still not entirely resolved by the model dynamics. Such resolutions, where the corresponding processes are neither well parameterized by traditional methods nor well resolved, are often referred to as the *grey zone* (Palmer, 2019b). Sometimes it is specified as the regime where the process is partially resolved, and partially subgrid-scale (Chow *et al.*, 2019; Kealy, 2019).

One obvious solution to the grey zone problem is to skip grey zone resolutions and directly use grid sizes where the model dynamics completely resolve the processes. Unfortunately, every parameterization has its own grey zone at a different scale so that resolutions beyond, e.g., the convective grey zone, are challenged by the turbulent grey zone, as indicated in Fig. 1 and Sec. 1.1.3. Hence Chow *et al.* (2019) refer to a gray continuum where the single grey zones of convection, topography and turbulence smoothly transition from one grey zone into the next and challenge mesoscale modeling of complex terrain. Due to the immense computational power necessary for only halving horizontal grid sizes (see Sec. 1.1.3), grey zones cannot be avoided overall. An alternative, commonly applied solution is to switch off traditional parameterizations. For example, deep convection is often not parameterized in models with grid lengths below a few kilometers, although they simulate only the largest convective updrafts explicitly (e.g., Baldauf *et al.*, 2011; Clark *et al.*, 2016). Consequently, the minimum size of the convective updrafts is determined by the grid length, and strong sensitivities to

resolution occur (Hanley *et al.*, 2015; Morrison, 2016b; Jeevanjee, 2017; Panosetti *et al.*, 2019).

These two solutions, which both bypass the actual problem, are not very satisfactory. Hence a lot of effort is currently devoted to address the grey zone and its problems properly. For instance, an international project is ongoing to synchronize model and observation data for the development of grey zone parameterizations for turbulence and convection (Tomassini *et al.*, 2019). One part of a solution can be the use of prognostic subgrid-scale variables or stochastic, temporally correlated variables to enable memory of the subgrid-scale variables (Rio *et al.*, 2019). Other methods use coupling with nearby grid boxes to relax the spatial scale separation, for instance, by using cellular automata (e.g., Bengtsson *et al.*, 2013). The uncertainty related to the probability density functions of the subgrid terms (Fig. 3) should also be appropriately quantified. This can be done by using stochastic parameterizations, which will also be used in this thesis. Details on such parameterizations are given in the following section.

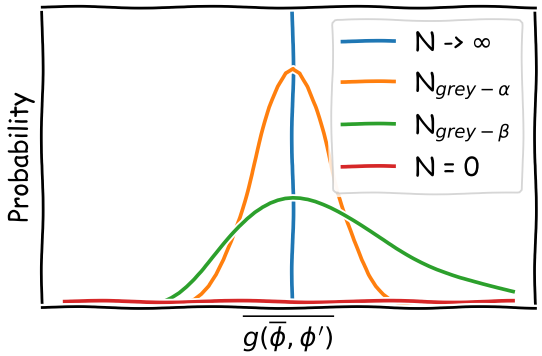


**Figure 3:** Probability density function of subgrid source terms are conceptually illustrated for the fully unresolved regime ($N = \infty$), the grey zone and the fully resolved regime ($N = 0$). At the higher resolution part of the grey zone ($N_{grey-\beta}$), the distributions are not necessarily Gaussian distributed.

### 1.2.2 *Stochastic parameterizations*

As the name implies, stochastic parameterizations are a form of parameterization that include a random component which can represent the variability of the subgrid-scale probability density functions as described above (Section 1.2.1) and illustrated in Fig. 3. In doing so, the impact from the subgrid-scales on the resolved scales is determined not only by the resolved variables but also by a random variable $\eta(\mathbf{x}, t)$, often a function of space and time:

$$\frac{\partial \overline{\phi}}{\partial t} \approx f(\overline{\phi}) + \underbrace{h_1(\overline{\phi}) + h_2(\overline{\phi}) \cdot \eta(\mathbf{x}, t)}_{\approx \overline{g(\overline{\phi}, \phi')}} .$$

See also equ. (1) for comparison. $h_1$ and $h_2$ represent general functions that need to be determined by the parameterization. Depending on the nature of $h_1$ and $h_2$, stochastic parameterizations can be multiplicative or additive. Stochastic parameterizations are mathematically well justified (Penland, 2003; Franzke *et al.*, 2015) and can be based on first principles using homogenization theory (Berner *et al.*, 2017). Indeed, several studies rigorously derive stochastic representations of subgrid effects in idealized models (e.g., Vissio and Lucarini, 2018; Chapron *et al.*, 2018; Demaeyer and Vannitsem, 2017, 2018).

For fully complex NWP or climate models, however, this is more challenging, if not impossible.

Nonetheless, such models can still benefit from more pragmatic stochastic parameterizations. First, stochastic parameterizations provide means to quantify model uncertainty, which is important for realistic ensemble spread and hence reliable models. Since many models are found to be under-dispersive (Berner *et al.*, 2017; Palmer, 2019b), stochastic parameterizations are often used to account for model error in a bulk way. A second advantage arises from the so-called *noise induced drift*. As illustrated in Fig. 4 overcoming nonlinear barriers will be more likely with stochastic forcing. Consequently, distributions of affected variables can change, biases be reduced, and the representation of extremes improved (Palmer, 2019b). Hence, stochastic parameterizations can affect not only the ensemble spread, but also the ensemble average of a simulation. A particular benefit arises for grey zones. The probability density function of subgrid terms can be appropriately represented, and memory terms or spatial correlations can further improve the coupling between subgrid and resolved terms.

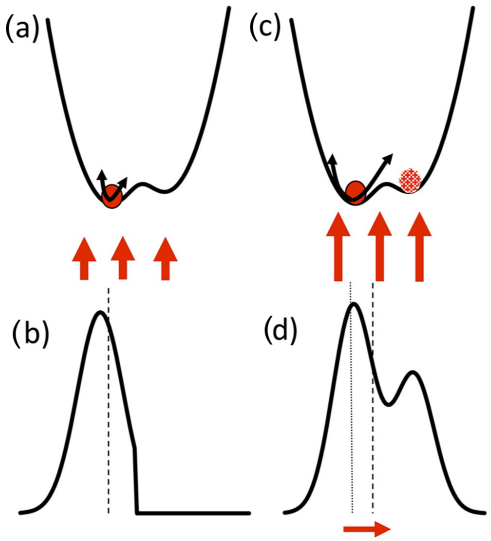


**Figure 4:** Conceptual illustration of noise-induced state transition by sufficiently strong noise. Too weak perturbations in the potential well in (a) prohibit a state transition, whereas the stronger noise in (c) enables state transitions. (b) and (d) illustrate the associated probability density functions. Figure from Berner *et al.* (2017), modified. ©American Meteorological Society. Used with permission.

For fully complex NWP or climate models, several stochastic parameterizations have been developed. Holistic approaches as the Stochastically Perturbed Parameterization Tendencies (SPPT) (Buizza *et al.*, 1999) and Stochastic Kinetic Energy Backscatter (SKEBS) (Shutts, 2005) approaches provide a bulk representation for many types of model error. In contrast, process-level stochastic perturbations directly represent model uncertainties at its known sources. Ollinaho *et al.* (2017) and Jankov *et al.* (2017, 2019) propose stochastic perturbations of uncertain model parameters (SPP). Christensen *et al.* (2017) update the SPPT scheme to treat different processes independently, which better fit observations compared to bulk approaches. Based on a statistical mechanics perspective of convection (Craig and Cohen, 2006), Plant and Craig (2008) introduce a physically motivated approach for deep convection and Sakradzija *et al.* (2016) for shallow convection. And Kober and Craig (2016) developed unique, physically based stochastic perturbations, the PSP scheme, to improve the coupling between boundary-layer turbulence and convective initiation in convection-permitting models. The PSP scheme will be addressed and revised in detail in Chapter 3 of this thesis.

## 1.3 RELEVANT PHYSICAL PROCESSES AND THEIR MODEL REPRESENTATION

As identified before (Section 1.1.3 and 1.1.2), a common model type for NWP today are km-scale models. Maximizing their skill is a current challenge for the NWP research community. One of the most effective ways to address this challenge is presumably to improve the initiation of convection by still subgrid-scale processes. As this will be addressed in detail in this thesis, we will now properly introduce atmospheric deep convection, its initiation, and how it is often represented in models. Then, we focus on the three most relevant processes for convective initiation based on their frequency, their different effects on convection and their insufficient respresentation in current models. These processes are boundary-layer turbulence, orographic convection and cold pools and are likely responsible for most deficiencies concerning convective initiation.

### 1.3.1 *Convection and its initiation*

Atmospheric deep convection represents the rapid formation of clouds in a conditionally unstable atmosphere. Such clouds extend throughout the depth of the troposphere and coincide with strong condensation and often heavy precipitation. In principle, the atmosphere is stably stratified, so that small displacements do not lead to an acceleration of rising air parcels. This is illustrated in Fig. 5a by the increase of environmental potential temperature with height throughout most of the troposphere. However, if air parcels are lifted beyond their lifting condensation level (LCL), clouds form due to condensation, latent heat is released, and the rising parcels become even warmer and less dense. At some point, called the level of free convection (LFC), the air will become less dense than its environment, and the lifted air will accelerate due to the continued release of latent heat (see Fig. 5a). This rising usually continues throughout the depth of the troposphere. Only at the tropopause, where temperatures start increasing with height, the acceleration ceases at the level of neutral buoyancy (LNB). This accelerated rising air is termed *atmospheric deep moist convection* or simply *convection*[6].

Typically, the possibility and strength of convection are characterized by the convective available potential energy, CAPE, and the convective inhibition, CIN. CAPE defines the energy that can be released due to evaporation and is given by the vertical integral of positive buoyancy $B$ (Markowski and Richardson, 2011):

$$\mathrm{CAPE} = \int_{LFC}^{LNB} B dz \approx g \int_{LFC}^{LNB} \frac{(\theta_{v,parcel} - \overline{\theta}_v)}{\overline{\theta}_v} dz.$$

The buoyancy is approximated in terms of the acceleration of gravity $g$ and the virtual potential temperatures of a horizontally uniform reference state $\overline{\theta}_v$ and a hypothetical, rising parcel $\theta_{v,parcel}$. In Fig. 5, we regard the reference temperature as characteristic for the environmental temperature with typical vertical profiles similar to the orange curve. The dry and moist adiabatic rising of a surface air parcel results in a typical vertical profile of $\theta_v$ similar to the blue curve. CAPE can then be interpreted as the

6 In the field of meteorology dry convection, i.e., buoyantly driven unsaturated thermals, or shallow moist convection are also often referred to by *convection*. In this thesis, however, we will use the term as an abbreviation for deep moist convection and specify other types of convection accordingly.

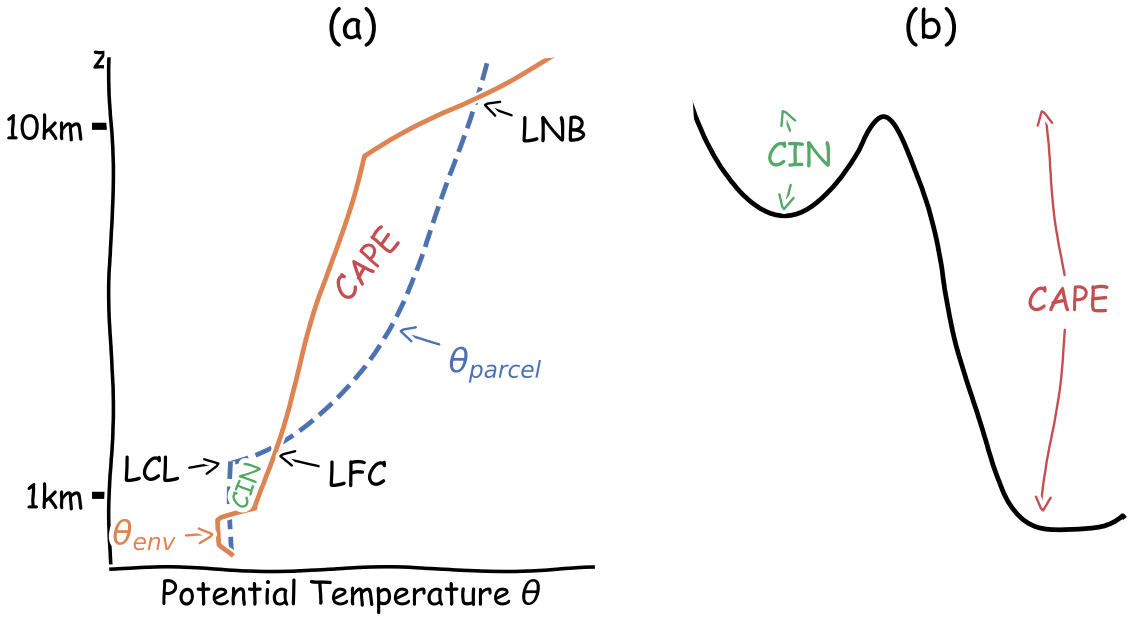


**Figure 5:** a) Conceptual illustration of environmental potential temperature $\theta_{env}$ and potential temperature of lifted air parcels $\theta_{parcel}$. The air parcel is lifted under dry adiabatic ascent (constant $\theta$) from the surface to its lifting condensation level (LCL) and under moist adiabatic ascent from the LCL upward. If the parcel's potential temperature is higher/lower than the environmental one, the parcel is positively/negatively buoyant and will accelerate/decelerate. b) Conceptual interpretation of CAPE and CIN in a potential well.

area between the parcel's and environmental temperature profiles (blue and orange curves) from LFC to LNB. Convection transforms this potential energy into kinetic energy, and vertical velocities of up to $w = \sqrt{2\text{CAPE}}$ can be reached. However, prior to the accelerated rising, the air first requires lifting to the LFC for which it usually needs to break through a layer of warmer, less dense air. This layer is called the convection inhibition layer, and the corresponding energy required for overcoming it is quantified by CIN. It is defined as the vertical integral of buoyancy from the surface up to the LFC (see also Fig. 5a):

$$\text{CIN} = -\int_{sfc}^{LFC} B dz \approx -g \int_{sfc}^{LFC} \frac{(\theta_{v,parcel} - \bar{\theta}_v)}{\bar{\theta}_v} dz.$$

When both CIN and CAPE $> 0$, the air parcels are stable to an infinitesimal perturbation but becomes unstable for sufficiently large perturbations, as illustrated in Fig. 5b by a potential well. Hence, for convection to occur, this CIN either has to be removed locally or overcome by forced lifting. We will refer to the process of overcoming CIN, either by local removal of CIN or by forced ascent, and reaching their level of free convection as *convective initiation*.

The described concepts are based on the so-called parcel theory. It assumes isolated air parcels that are distinct from their environmental air and do not interact with it. While some observations are fairly consistent with parcel theory (Trapp, 2013), it strongly simplifies convective processes. Perturbation pressure gradients, entrainment/detrainment, hydrometeor loading, latent heating from freezing water droplets or

compensating environmental subsidence are neglected (Markowski and Richardson, 2011). A major implication is that buoyant acceleration is overestimated as persistent plumes are often necessary to sustain the buoyancy of detraining air (Trapp, 2013).

Convective initiation can occur due to numerous processes and their interactions. Surface heating can locally remove CIN or produce unstable, rising air parcels or turbulent eddies with sufficient buoyancy or kinetic energy to overcome CIN (see Sec. 1.3.2). Furthermore, several aspects of orographically induced flow can result in lifting to the LFC (see Sec. 1.3.3). A third mechanism that is actively investigated in many studies is the secondary convective initiation by cold pools (Sec. 1.3.4). Further processes include synoptically driven frontal systems, convergence lines, horizontal convective rolls or mesoscale circulations like land see breezes.

As already described in Section 1.1.3, the representation of convection in models strongly depends on the used grid sizes. NWP models with grid sizes of $\mathcal{O}(10\,km) - \mathcal{O}(100\,km)$ (e.g., global NWP models, see Sec. 1.1.3) are not able to resolve these deep convective overturning circulations. Instead, parameterizations are used to represent their effect on the resolved variables. Most schemes apply a mass flux approach introduced by Arakawa and Schubert (1974), where fast convective mass fluxes are assumed to balance large scale and radiative forcing (quasi-equilibrium assumption). A wide variety of different approaches and modifications exist and are still being developed (Rio *et al.*, 2019). Current development focuses explicitly on improving the transition from shallow to deep convection, unifying shallow and deep convection schemes, including convective memory and stochastic schemes to account for cloud size distributions (Rio *et al.*, 2019).

Nonetheless, due to the many underlying approximations and systematic biases of convection parameterizations (e.g., Gentine *et al.*, 2018) models that can explicitly resolve deep convective overturning are highly desirable (e.g., Clark *et al.*, 2016). As described in Section 1.1.3, such convection-permitting models require grid sizes of few kilometers or less and are now feasible for some applications like regional weather prediction and will likely become important also for other applications (see Section 1.1.3). Fig. 6 visualizes such resolved clouds in a convection-permitting model with grid sizes of $2\,km$ and illustrates how crude such cloud simulations still are. Many grey zone problems are consequently highly relevant for these models (see Sec. 1.2.1) and deficits in precipitation forecasts exists.

Even though deep convective circulations can be explicitly simulated, several tightly coupled processes cannot. Shallow cumulus clouds are usually still parameterized, which results in an often inaccurate transition from shallow to congestus and deep convection (Baldauf *et al.*, 2011; Rio *et al.*, 2019). Also, such models still have to parameterize turbulence, which results in an unsatisfactory turbulent exchange of environmental air with the cloud, i.e., entrainment and detrainment (de Rooy *et al.*, 2013; Rio *et al.*, 2019). Tang and Kirshbaum (2020), for instance, show resolution sensitivities of convective initiation even in hectometer scale models, which are partially related to turbulent mixing. On a more fundamental level, vertical acceleration itself depends on the horizontal extent of the forcing. For convective scale processes in km-scale models, this horizontal extent is constrained by the horizontal resolution of the model, and vertical velocities are reduced (Weisman *et al.*, 1997; Pauluis and Garner, 2006; Morrison *et al.*, 2015; Morri-

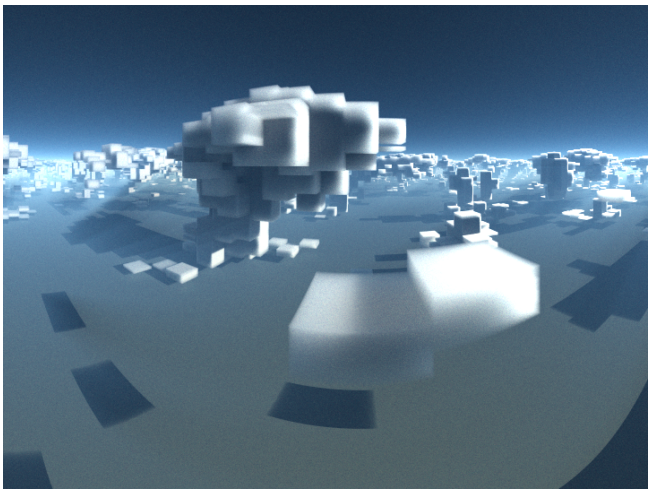

**Figure 6:** This figure visualizes resolved convective clouds in a convection-permitting model in three dimensions. The white cubes correspond to cloudy grid boxes in the model. Simulations were computed with the EULAG model using 2 km grid sizes. The figure ist taken from Mayer (2018), with courtesy of B. Mayer.

son, 2016a; Jeevanjee and Romps, 2016; Jeevanjee, 2017). This quite certainly will affect the initiation and behavior of convection. Furthermore, processes for overcoming the convective inhibition are often still subgrid-scale and hinder the initiation of convection. Such processes include boundary-layer turbulence, orographically induced lifting or cold pools. We will address the coupling of these processes with convective initiation within convection-permitting models throughout this thesis.

These problems consequently impact the quality of precipitation forecasts. Hanley *et al.* (2015), for instance, find a strong sensitivity of precipitation cell sizes, intensity or initiation time to the mixing length parameter of the subgrid turbulence scheme. Biases in the diurnal cycle of convection have also been identified by Baldauf *et al.* (2011); Clark *et al.* (2016); Rasp *et al.* (2018b), and a lack of persistent organization into the night was identified by Rasp *et al.* (2018b).

### 1.3.2 *Boundary layer turbulence*

The planetary boundary layer (PBL) is the lowest part of the troposphere where the impact of the earth's surface on atmospheric momentum, heat and moisture is crucial. The upper height of the PBL varies from 100 m to a few kilometers. The dominant transport of surface properties, like momentum, heat or moisture, occurs via 3D turbulence, which facilitates a fast response of the PBL to surface changes. As a consequence, the summer-time boundary layer typically follows a diurnal evolution forced by radiation, as depicted in Fig. 7. After sunrise, surface heating induces turbulent sensible heat fluxes, which enable vertical mixing within the so-called mixed layer (D). Overshooting thermals penetrate the stably stratified atmosphere above until their negative buoyancy compensates their kinetic energy (D, E). This layer is referred to as the entrainment zone and separates the boundary layer from the free atmosphere. Then, the mixing layer grows with continued surface heating (D, E, A). The turbulent eddies are mostly generated by buoyancy and can extend throughout the whole depth of the mixed layer. After sunset, outgoing longwave radiation cools the surface and stabilizes the lower boundary layer (B, C). In this stable boundary layer, buoyancy-driven turbulence is suppressed, but intermittent turbulence driven by shear can still occur.

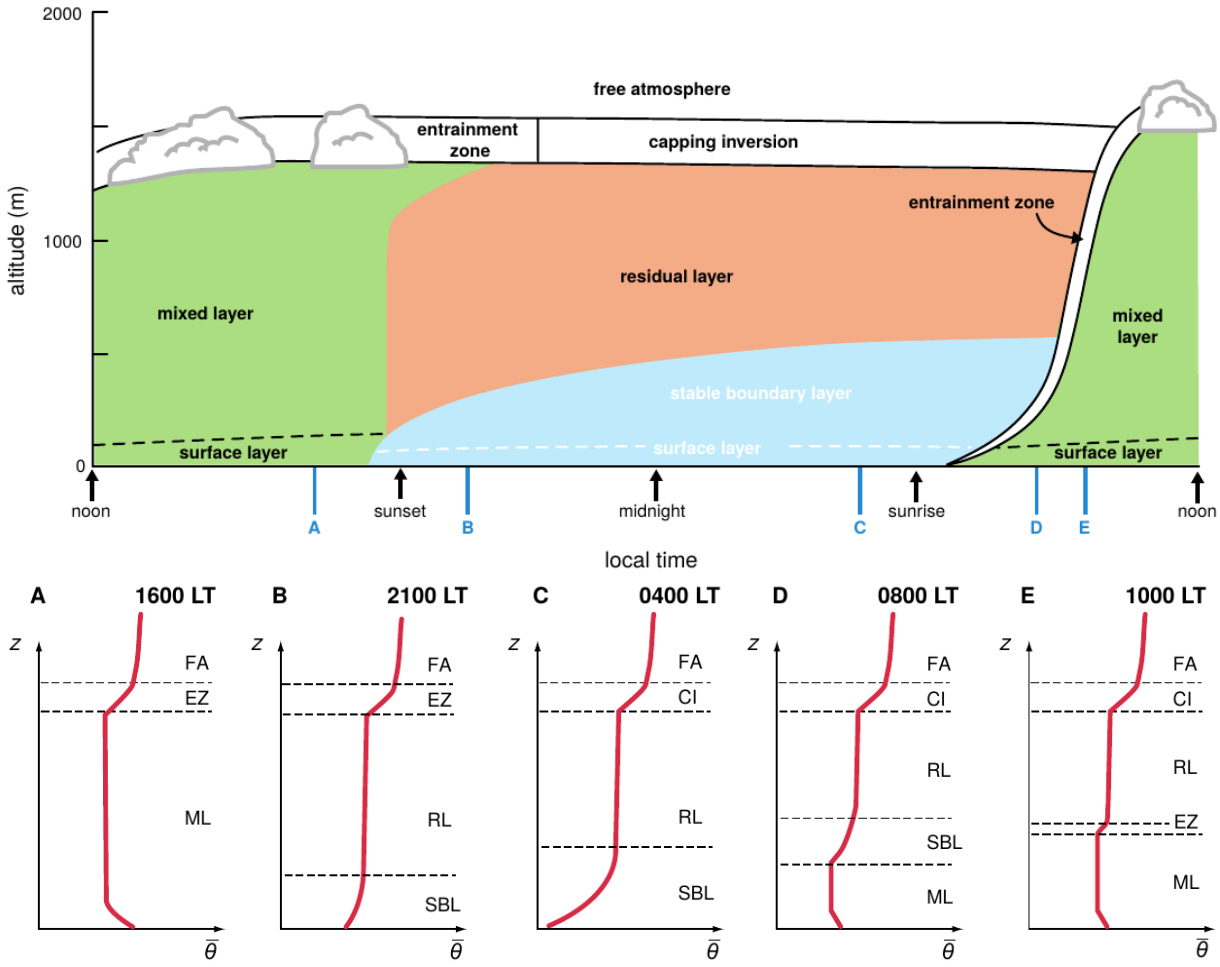


**Figure 7:** A typical diurnal evolution of the planetary boundary layer in summer time conditions. The top figure displays the schematic evolution while the bottom figures show vertical profiles at selected times. From Markowski and Richardson (2011).

Turbulence can be formally quantified by the *turbulent kinetic energy* (TKE) $\overline{e}$. It is defined as the Reynolds averaged, mass-specific kinetic energy of perturbation velocities $u'$, $v'$, $w'$:

$$\overline{e} = \frac{1}{2}\left(\overline{u'^2} + \overline{v'^2} + \overline{w'^2}\right)$$

TKE is not a conserved quantity, and several sources and sinks exist. On the molecular scale, viscous dissipation transfers turbulent energy to internal energy at the molecular scale and hence always decreases TKE. It is, however, only dominant for scales smaller than the Kolmogorov length scale, which is in the order of millimeters for the atmosphere (Vallis, 2017). The most frequent drivers for TKE in the boundary layer are given by buoyancy and shear. The former is based on vertical heat fluxes that produce TKE under unstably stratified situations and decrease TKE in stable situations. The latter, usually dominated by vertical shear of horizontal winds, can generate TKE very locally. In stable situations, e.g., in a nighttime PBL, TKE can only be produced if the shear production dominates over buoyant suppression. See Markowski and Richardson (2011) for more details or Stull (1988) and Wyngaard (2010) for a comprehensive treatment.

As described in the previous section, the boundary layer is closely linked to convective initiation. The depth of the boundary layer, in combination with the temperature difference in the entrainment zone, determine the convective inhibition (see Fig. 5). Furthermore, if a parcel's buoyancy is sufficiently large, it can reach the LFC and transform into a deep convective updraft.

So how can chaotic turbulence be represented by NWP models with grid sizes larger than the typical scale of the eddies? By Reynolds-averaging the primitive equations, resolved scale model equations can, in principle, be derived from first principles. The arising equations constitute prognostic equations for Reynolds averaged quantities. For the x-component of the Navier-Stokes equation under the assumption of constant density $\rho$, we obtain the following equation:

$$\frac{\partial \overline{u}}{\partial t} + (\overline{\mathbf{v}} \cdot \nabla)\overline{u} = -\frac{1}{\rho}\frac{\partial \overline{p}}{\partial x} - \nabla \cdot (\overline{\mathbf{v}'u'}), \tag{2}$$

$$\text{with } \nabla \cdot (\overline{\mathbf{v}'u'}) = \frac{\partial}{\partial x}\overline{u'u'} + \frac{\partial}{\partial y}\overline{v'u'} + \frac{\partial}{\partial z}\overline{w'u'}$$

(Vallis, 2017). The last term in (2), the *Reynolds stress term*, describes the influence of the subgrid-scale eddies on the resolved term. The arising subgrid terms for all averaged equations can be categorized into *turbulent kinematic momentum fluxes*, consisting only of velocity components, *turblent kinematic heat fluxes* (e.g., $\overline{w'\theta'}$) and *turbulent kinematic moisture fluxes* (e.g., $\overline{w'r_v'}$). As these terms cannot be solved explicitly by the model, they need to be specified using resolved, i.e., Reynolds averaged variables. A variety of approaches exist to determine these subgrid terms (Stensrud, 2007). Usually, such PBL schemes assume horizontally homogenous situations and, thus, only consider vertical flux terms. Hence, such schemes are one-dimensional turbulence parameterizations. Local schemes determine the turbulent fluxes only by using local flow characteristics. Non-local schemes, in contrast, often include characteristics of the whole boundary layer.

A common local approach applies the so-called *K-theory*. Doing so, these terms are viewed in analogy to molecular viscosity, so they are approximated by the resolved scale gradients, e.g., $\overline{u'w'} \approx -K_m \frac{\partial \overline{u}}{\partial z}$. $K_m$ is denoted as the *eddy viscosity* and needs to be specified. Often a characteristic mixing length $l$ is used to determine $K_m$. Higher-order schemes include prognostic equations also for eddy terms. Then, however, third or even higher-order eddy terms arise which need a parameterization. One example is a so-called 1.5 TKE closures, where a prognostic equation for the TKE is used to specify the eddy correlation terms. Common schemes that combine local and non-local behavior are eddy diffusivity/mass flux schemes (EDMF) (Siebesma *et al.*, 2007) and are now also being developed as unified parameterization for turbulence and shallow or deep convection (Tan *et al.*, 2018) or with stochastic components (Sakradzija *et al.*, 2016).

As horizontal grid sizes reach the scale of the boundary-layer height, the assumptions of the Reynolds averaged equations break down.[7] If grid sizes become so small that

7 As argued in Wyngaard (2010), Reynolds averaging should at best refer to an ensemble average, i.e., the average over many realizations under the same larger-scale conditions. Due to the often practical limitations of an ensemble average, it can be approximated by temporal or spatial averages under certain

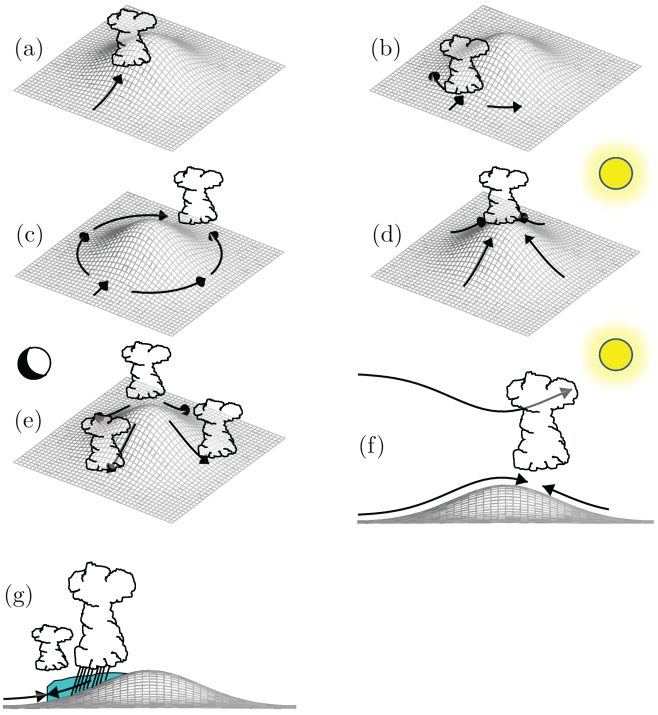


**Figure 8:** Different mechanisms of convection triggering by orography. See text for details. From Kirshbaum *et al.* (2018).

the ratio of the typical eddy length $l$ and grid size $\Delta x$ is large, i.e., $\frac{l}{\Delta x} \gg 1$, most eddies can be explicitly resolved by the model and only the smallest scales have to be parameterized. Simulations in this regime are referred to as *Large eddy simulations* (LES), and established 3D turbulence parameterization schemes exist. If, however, $\frac{l}{\Delta x} \approx 1$ neither the LES nor the Reynolds averaging approaches are valid. Wyngaard (2004) termed this regime as the *terra incognita*, the grey zone of turbulence parameterizations. Suitable schemes are actively being developed (see Chow *et al.*, 2019, for a review) and often blend between the 3D LES and 1D PBL approach in a scale adaptive way (Hanley *et al.*, 2019; Kealy *et al.*, 2019; Kealy, 2019).

### 1.3.3 *Orographic convection and the role of subgrid-scale orography*

#### *Orographic convection*

Various orography related processes have been proposed for the initiation of convection. In general, orographically induced convection can be categorized into mechanically and thermally driven convection and their interactions. Most relevant processes are illustrated in Fig. 8.

For mechanically induced convective initiation by circular orography, the non-dimensional mountain height $M$ categorizes two different flow types:

$$M = \frac{Nh_m}{U},$$

conditions. These conditions arise from ergodic theory: under stationary or homogenous situations, the temporal or spatial average converges to the ensemble mean. In real-world phenomena, only quasi-stationary or quasi-homogeneous situations exist, and the question arises how long/far we must average to yield a temporal/spatial average that sufficiently represents the ensemble average. Averages over grid sizes of 10-100 km are likely sufficiently accurate.

with the Brunt-Väisälä freqeuncy $N$, the mountain height $h_m$ and a characteristic horizontal wind speed $U$. For $M < 1$, the horizontal flow contains sufficient kinetic energy to overcome the height barrier of the mountain, and if the lifted air reaches its LFC, convection occurs (Fig. 8a). We will refer to this mechanism as *mechanical lifting* and specifically focus on it in this thesis. For $M > 1$, on the other hand, the kinetic energy is insufficient, and the flow is *blocked* by the mountain. Several possibilities for initiating convection occur. Flow deceleration by upstream blocking may result in convection on the mountain luv (Fig. 8b). Another possibility is the lee-side convergence with subsequent convection, as illustrated in Fig. 8c. Additionally, mountain generated gravity waves and hydraulic jumps can result in further convective initiation.

Thermally driven circulations can result from surface heating or cooling. During the day, diabatic surface heating results in higher potential temperatures at mountain tops (elevated heating) compared to the valley. Thus, on similar height levels, temperature gradients exist that induce a baroclinic circulation resulting in up-slope winds, as illustrated in Fig. 9. A subsequent convergence at the mountain top can then trigger convection (see also Fig. 8d). During nighttime katabatic, down-slope winds can occur and lead to mountain base convective initiation (Fig. 8e).

Interactions between up-slope winds and gravity waves with background wind situations (Fig. 8f) or cold pools (Fig. 8g) can further induce convection. The specific relevance and technicalities of each process are still the topic of current research. Further details can be found in the recent review by Kirshbaum *et al.* (2018) on orographic convection or in the more general review by Houze (2012) on orographic precipitation.

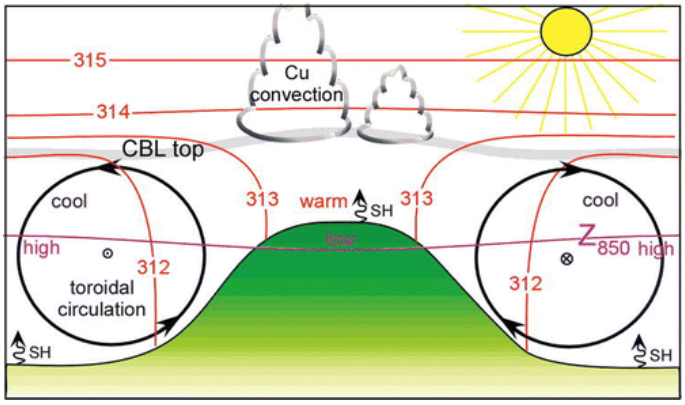


**Figure 9:** Illustration of thermally induced orographic circulations. The warming of elevated surfaces results in higher temperatures at the mountain compared to the same height above the valley, as illustrated by the red isolines of potential temperature. This horizontal temperature gradient induces a baroclinic circulation represented by the black circles. If the upward branch of the circulation reaches the LFC, convection can occur on top of the mountain. From Geerts *et al.* (2008). ©American Meteorological Society. Used with permission.

*The role of subgrid-scale orography*

While the specific relevance and details of single processes may still be an unsettled topic, the overall relevance of orography for convection is undeniable. In numerical

models, however, the effects of orography are, in general, only included if the model grid sufficiently resolves orography. With decreasing grid sizes of numerical models, more and more of the fractal orography is resolved by the model grid, enabling an explicit representation of these processes. In fact, better resolving orography is often mentioned as a dominant benefit of increased resolutions (Wagner *et al.*, 2014; Kealy, 2019).

Nonetheless, it is also possible to include the effects of subgrid-scale orography (SSO) to some extent via SSO-parameterizations. The effect of non-local orographic drag is represented in parameterizations for orographic gravity waves (Lott and Miller, 1997). Furthermore, the more local effect is accounted for in the specification of the roughness length. The recent development of parameter perturbations even allows for uncertainty in the roughness length (Ruckstuhl and Janjić, 2020). Current developments further aim to properly include SSO in boundary-layer turbulence parameterizations (Rotach *et al.*, 2015). The influence of SSO on convection, however, is otherwise not parameterized in models.

This consequently raises the question of whether SSO, i.e., small orographic scales, is, in fact, necessary for convection. Several studies have addressed this question. While Kirshbaum *et al.* (2007a) find that scales smaller than few kilometers are not dominant for convective initiation when a range of orographic scales exist, various studies emphasize the role of small scale orographic features using analytical models, idealized simulations or case studies. Panosetti *et al.* (2016) find differences in thermally induced orographic convection between idealized LES and CRM simulations. Kirshbaum *et al.* (2007b,a) also show that small, single scale orographic features can dictate the spacing of convective bands, and Langhans *et al.* (2011) conclude that small scale orography contributes significantly to convective initiation. By evaluating the effect of smoothed terrain on convection in high-resolution simulations, Schneider *et al.* (2018) find complex interactions between background flow and orographic scales. These diverse results illustrate the complexity of the problem and that the relevance of small orographic scales for convection is still uncertain.

Despite these ambiguous conclusions for small scale orography, and hence SSO, it could be worth to consider its effect for convective initiation in a parameterization. In general, orography can act as an intrinsic source of predictability for otherwise often unpredictable summertime convection (Bachmann *et al.*, 2019, 2020). Also, Kirshbaum *et al.* (2018) recognizes the lack of SSO in convection parameterizations as a "longstanding deficiency". The authors further emphasize that, within the "mountain 'grey zone', where a smoothed profile of a mountain range is explicitly resolved, but important smaller-scale forcing is not", (Kirshbaum *et al.*, 2018, p.19) the development of scale aware parameterizations is required to include convective initiation by SSO.

### 1.3.4 *Cold pools*

Cold pools are volumes of negatively buoyant air that originate from precipitating downdrafts. Evaporation of precipitation in the sub-cloud layer, combined with the weight of the condensed water, creates negative buoyancy and thereby accelerates the downdraft. When these cold, moist air masses hit the surface, they spread in circular-

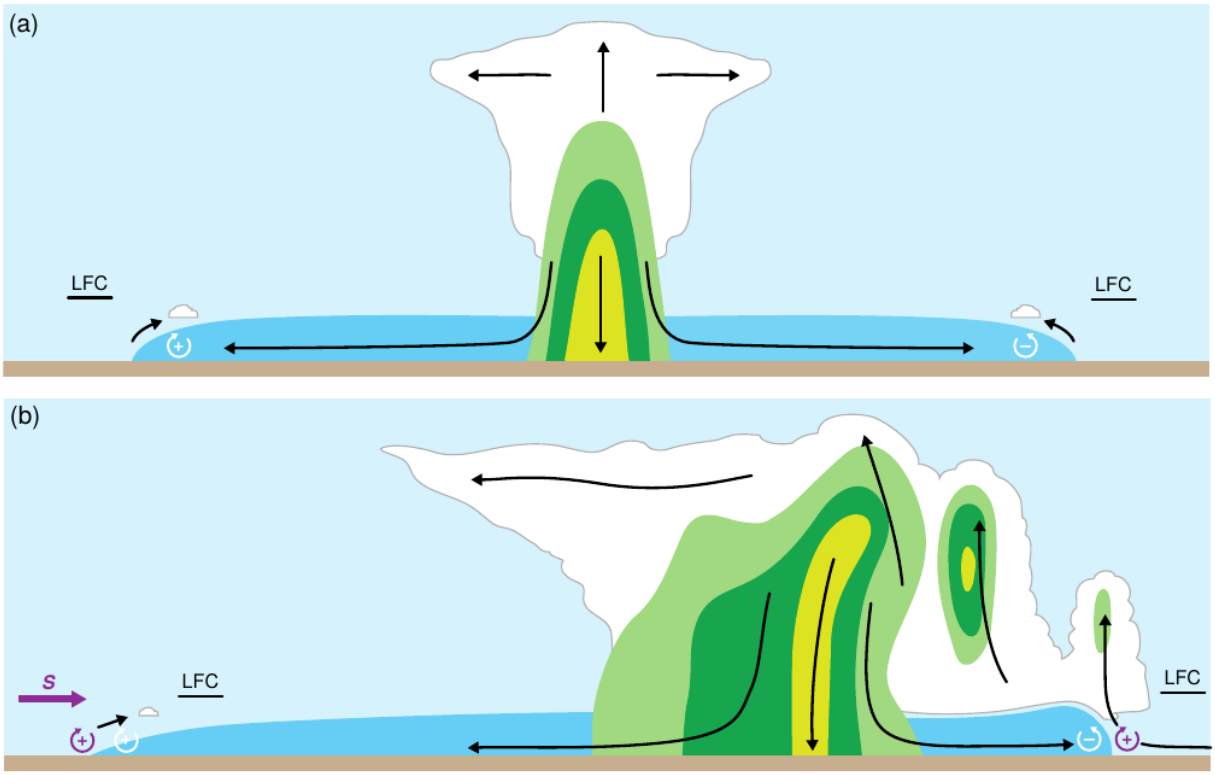


**Figure 10:** Schematic illustration of cold pool generatation by convective downdrafts (a) and maintenance of convective cells by secondary convective initiation via mechanical forcing (b). Shear-driven horizontal vorticity is displayed in purple, buoyancy related horizontal vorticity in white. From Markowski and Richardson (2011).

like shapes, as illustrated in Fig. 10a. Cold pools can be regarded as density currents, which are well studied fluid dynamical objects. One well-established framework provides the possibility to estimate the propagation speed of density currents as $U = \sqrt{BH}$, where $B$ represents the buoyancy anomaly of the cold pool and $H$ the height of the anomaly (see, e.g., von Kármán, 1940; Benjamin, 1968; Bryan and Rotunno, 2014b,a). In the atmosphere, cold pools actively interact with the surface (Peters and Hohenegger, 2017; Grant and van den Heever, 2018; Gentine *et al.*, 2018) and are lead by a gust front, where lifting occurs that fosters the initiation of new convection (Fig. 10b).

In recent decades, three major mechanisms have been proposed as explanations for this secondary convective initiation. First, Rotunno *et al.* (1988) and Weisman and Rotunno (2004) proposed that horizontal vorticity linked to cold pool induced buoyancy gradients at the gust front interacts with the vorticity generated by the background wind shear to yield approximately vertical updrafts. This pattern of convergence and ascent is optimal for initiating new convection and thereby enables the reinforcement and propagation of squall lines (Rotunno *et al.*, 1988; Weisman and Rotunno, 2004; Markowski and Richardson, 2011). We will refer to this mechanism as mechanical lifting. Second, the collision of cold pools has been identified as a major contributor for initiating new convection (Feng *et al.*, 2015; Cafaro and Rooney, 2018; Haerter *et al.*, 2018; Torri and Kuang, 2019). Collisions of two cold pools can cause a superposition of their updraft velocity resulting in stronger updrafts. Collisions of three cold pools can

trap non-cold pool air in between the colliding fronts, which results in even stronger vertical motion (Haerter *et al.*, 2018). Third, Tompkins (2001) proposed a mechanism based on buoyancy-driven initiation: decaying cold pools accumulate moisture from the perished precipitation downdrafts at their boundaries. This increased moisture can compensate the effect of cold temperatures on buoyancy (especially for "old" cold pools, where entrainment has depleted the cold air) and thereby reduce the convective inhibition, which supports the initiation of new convection. Torri *et al.* (2015) investigated the relevance of the mechanical and thermodynamic mechanisms in idealized studies and found that, initially, mechanical lifting drives the vertical motion, whereas higher in the boundary layer, moisture seems to be relevant for vertical acceleration. Independent of the specific mechanism, the effect of cold pools is to trigger convection in the vicinity of already existing convection, leading to the organization of convective cells and a more clustered precipitation pattern. Cold pools may also influence the diurnal cycle of convection since they are relevant triggering mechanisms, only after initial precipitation events have occurred. Especially in the late afternoon/evening, when other mechanisms like surface heating are no longer strong, cold pool triggering may dominate.

In coarser weather and climate models, cold pools are often not resolved. Yet only recently are the effects of cold pools being considered in convection parameterizations (e.g., in Grandpeix and Lafore, 2010; Grandpeix *et al.*, 2010), often to improve convective memory and organization (Rio *et al.*, 2019). With km-scale grid sizes, models can simulate cold pools explicitly. Cold pool properties, however, are highly sensitive to changes in model resolution, suggesting that these models are not able to simulate cold pools accurately. Grid sizes of below $100\,m$ are expected to be necessary for accurately resolving the fine-scale structures of cold pool gust fronts (Grant and van den Heever, 2016). How to best tackle this cold pool grey zone is still an open question. A coexistence of resolved and parameterized cold pools may be useful for the coarser grid sizes, whereas alternative approaches might directly improve the resolved cold pools.

## 1.4 SUMMARY AND RESEARCH GOALS

A large part of precipitation and associated hazards comes from deep convective clouds which develop due to accelerated rising air by condensation. To reduce associated damages, early and reliable predictions are crucial. Predicting such deep convection, however, has been a long-standing challenge. In the last decade, km-scale models, which enable the explicit simulation of such deep convection (i.e., convection-permitting models), have become more and more accessible for operational weather prediction (Baldauf *et al.*, 2011; Clark *et al.*, 2016) and even climate projections (Leutwyler *et al.*, 2017; Stevens *et al.*, 2019a), and they are expected to prevail for the next decades (Palmer, 2019a). Consequently, convection parameterizations - containing many approximations and systematic biases (Gentine *et al.*, 2018) - are not required any more, and convection forecasts were significantly improved (Baldauf *et al.*, 2011; Clark *et al.*, 2016). Nonetheless, not all deficits concerning convective precipitation have been solved just by turning from parameterized convection to convection-permitting simulations. Most importantly, capturing the diurnal cycle of convection is still problematic (Baldauf *et al.*, 2011; Clark *et al.*, 2016; Hanley *et al.*, 2019) and a lack of persistent organization into the night

was identified by Rasp *et al.* (2018b). These inconsistencies become particularly relevant when synoptic forcing is weak, and local mechanisms are the main driver for overcoming convective inhibition (Keil *et al.*, 2014).

The underlying causes for these deficits can mostly be ascribed to the grey zone problem: grid sizes of convection-permitting models only allow the explicit simulation of large deep convective clouds; smaller convective systems and many processes, that are tightly coupled to convection, can still not sufficiently be resolved. These processes include evaporation and condensation, droplet formation, entrainment or detrainment, but, most importantly, the initiation of convection (Baldauf *et al.*, 2011; Yano *et al.*, 2018; Kober and Craig, 2016): for the formation of deep convective clouds to be kicked off, air parcels need to overcome the convective inhibition, which often requires local triggering processes related to boundary-layer processes. Currently, the most efficient way to improve convection is likely by improving the reprentation of processes that are relevant for convective intiation.

Hence, in this this thesis, we aim to understand how such triggering processes are insufficiently accounted for in convection-permitting models and to improve their representation accordingly. Based on their frequency, their distinct influence on convection and their insufficient representation in current models, we choose the three processes that are most promising to benefit convective initiation. These processes are boundary-layer turbulence, subgrid-scale orographic lifting and cold pools.

BOUNDARY LAYER TURBULENCE Boundary layer turbulence is an omnipresent phenomenon in summer-time boundary layers, occurs in a variety of situations and consequently contributes substantially to total convective initiation. It is specifically important during summer-time situations when synoptic, resolved scale forcing mechanisms are weak. Boundary layer turbulence is a subgrid-scale process in km-scale models and only represented approximately via parameterizations. Km-scale grid sizes specifically challenge traditional turbulence parameterizations as they approach the grey zone of turbulence. To improve its representation in km-scale models, we consider one of the first process-level stochastic parameterizations, the aforementioned PSP scheme (Kober and Craig, 2016), which introduces stochastic perturbations to account for the missing subgrid-scale variability of turbulence. We substantially revise the PSP scheme to improve its physical consistency and to eliminate some undesired effects. For example, previously ineffective vertical velocity perturbations are improved to be more persistent by also perturbing horizontal wind components in a 3d non-divergent way. The revised PSP scheme, PSP2, is then applied on a 10 day period to confirm its impact.

SUBGRID-SCALE OROGRAPHIC LIFTING The next most relevant process is thought to be orographically driven convective initiation. It is limited to orographic regions, yet many parts of Germany, Europe or the whole earth are indeed subject to complex terrain. Small scale orography, i.e., subgrid-scale orography, is not explicitly resolved, and its effect can only be approximated in the form of parameterizations. As explained in Section 1.3.3, however, it is not yet clear how relevant small scale orography, i.e., subgrid-scale orography in km-scale models, is for convective initiation. To contribute

to this unanswered research question, we develop stochastic perturbations similar to the PSP scheme, the SSOSP, to account for mechanical lifting by subgrid-scale orography.

COLD POOLS The largest part of this thesis addresses cold pools. Cold pools are a more local and time-dependent phenomenon, and proper observations are still missing to quantify their frequency. Cold pools interact strongly with convection itself and are considered particularly for their role in organizing convection and for producing persistent precipitation into the evening. While km-scale models indeed simulate cold pools themselves, many of their characteristics are still sensitive to the resolution, and their ability to trigger new convection is questionable. In the first cold pool project, we investigate in detail which aspects of cold pool driven convective initiation convection-permitting models are struggling with. We analyze the resolution sensitivity of cold pools, cold pool gust fronts and related convective initiation in high-resolution simulations. We further apply a linear causal analysis to better understand, how model resolution impacts cold pool driven convective initiation via different indirect effects. In the second cold pool project, we apply the gained understanding to develop a deterministic cold pool parameterization, the CPP scheme, where we strengthen cold pool gust fronts to improve cold pool driven convective initiation. This CPP scheme is then also applied on ten days to evaluate its impact, and its combination with PSP2 is tested.

We summarize the arising **research questions** as follows:

**Chapter 3** How can we improve the physical consistency of the PSP scheme to represent the influence of **subgrid-scale turbulence** on convective initiation better?

**Chapter 4** How can we represent the effects of **subgrid-scale orography** for convective initiation and how relevant is subgrid-scale orography for convective initiation in km-scale models?

**Chapter 5** What are convection-permitting models missing with regard to **cold pool driven convective initiation**?

**Chapter 6** How can we improve the representation of **cold pool driven convective initiation** in km-scale models and how does this affect precipitation?

By adressing these research questions, we provide new understanding on the representation of convective initiation. This will help to guide future research. We further provide innovative approaches to improve the representation of convective initiation in numerical models at grey zone resolutions. The most beneficial aspects are, first, the better representation of the processes which will likely improve future forecasts; second, the quite unique combination of process-based design with stochastic components, which allows a better quantifcation of model uncertainty at its source. This is a crucial, but rarely considered, way to get more reliable forecasts. Furthermore, the schemes can partially adjust to different model resolutions. Also, tuning, to compensate for insufficiently represented processes, can be reduced.

## 1.5 OUTLINE

This thesis is structured as follows. The COSMO model and setup, which is used throughout this thesis except for Chapter 5, is described in **Chapter 2**, as well as related verification metrics and used radar observations. Boundary layer turbulence is addressed in **Chapter 3**, where we comprehensively revise this PSP scheme and subsequently evaluate a new version on a 10-day test period. The representation of subgrid-scale orographic lifting for convective initiation is developed and minimally evaluated in **Chapter 4**. Cold pool driven convective initiation is addressed in Chapter 5 and 6. A comprehensive evaluation of existing deficits in **Chapter 5** facilitates insight on how cold pool driven convective initiation is insufficiently represented. The gained insight is then used to develop cold pool related perturbations in **Chapter 6** to reduce corresponding deficits. We close with a summary and conclusions in **Chapter 7**.

## 1.6 PUBLICATIONS

Parts of this dissertation are published as the following two publications:

(a) ***Hirt, M.**, Rasp, S., Blahak, U., & Craig, G. (2019). Stochastic parameterization of processes leading to convective initiation in kilometre-scale models. Monthly Weather Review, 147, 3917–3934, doi: 10.1175/MWR-D-19-0060.1. ©American Meteorological Society. Used with permission.*

This publication addresses the modifications to the PSP scheme, (i.e., Chapter 3 in this thesis) and the development of the SSO scheme (Chapter 4 here). Parts of Chapter 1, 2 and 7 are also included in this publication.

(b) ***Hirt, M.**, Craig, G. C., Schäfer, S., Savre, J., & Heinze, R. (2020). Cold pool driven convective initiation: using causal graph analysis to determine what convection-permitting models are missing. Quarterly Journal of the Royal Meteorological Society, doi: 10.1002/qj.3788*

This second manuscript addresses cold pool driven convective initiation and its sensitivity to model resolution in high-resolution simulations. Doing so, we identify potential deficits in convection-permitting models. This publication corresponds to Chapter 5 in this thesis. Parts of Chapter 1 and 7 are also included in this publication.

# 2

# COSMO: MODEL SETUP, SIMULATION PERIOD AND VERIFICATION

For chapters 3, 4 and 6, we aim to improve convective precipitation forecasts by developing better representations of boundary-layer processes for state-of-the-art, convection-permitting models. With this objective, we choose the COSMO model in a convection-pertmitting setup as it is used for the operational weather prediction by the German Weather Service. To test the impact of the parameterization schemes and their adaptability to different flow situations, we select a ten day time period with active convection and different synoptic situations. The used domain over Germany also contains orographic regions in the Southern part of the domain (Alps) and flat terrain in Northern Germany. This enables the better evaluation of the subgrid-scale orography scheme in different orographic situations. To evaluate the influence of the schemes on precipitation, we compare the simulation results to radar observations of precipitation, which have a high spatial coverage compared to other types of precipitation measurements. This comparison entails the computation of verification metrics, which allow meaningful insight in intermittent precipitation fields. Further details on these aspects are given below. Simulations and methods used for Chapter 5 will be explained then.

## 2.1 MODEL AND SIMULATION SETUP

Most simulations (except for those used in Chapter 5) are computed with the COSMO (COnsortium for Small-scale MOdeling) model (Baldauf *et al.*, 2011), version 5.4g, in a convection-permitting setup with $\Delta x = 0.025°$, roughly 2.8 km, and a time step of 25 s. The setup uses 50 vertically stretched model layers ranging from 10 m above ground to 22 km above mean sea level. A staggered grid (Arakawa-C/Lorenz) is used with terrain following coordinates (Schättler *et al.*, 2016). As a consequence, a proper transformation between model level and height depends on the horizontal coordinates. For comparability, we will give approximate, domain averaged transformations when necessary. Shallow convection is parameterized using the Tiedtke scheme. Details on parameterizations can be found in Doms *et al.* (2011). The setup follows the operational COSMO-DE setup with 461 by 421 grid points centered over Germany at 10°E and 50°N.

The only major deviation from the current operational setup is a change of the tuning parameter `tur_len` in the boundary-layer scheme to 500 m. For the operational COSMO version, a decision was made to decrease `tur_len` to 150 m in order to reduce a lack of summer-time convection (Baldauf *et al.*, 2011). This, however, comes at the expense of temperature biases, which are widely observed in different studies (Baldauf *et al.*, 2011; Schraff *et al.*, 2016; Leutwyler *et al.*, 2017; Necker *et al.*, 2018). Baldauf *et al.* (2011) argue that "the improved forecasts of deep convection outweigh those disadvantages" (p. 3902,

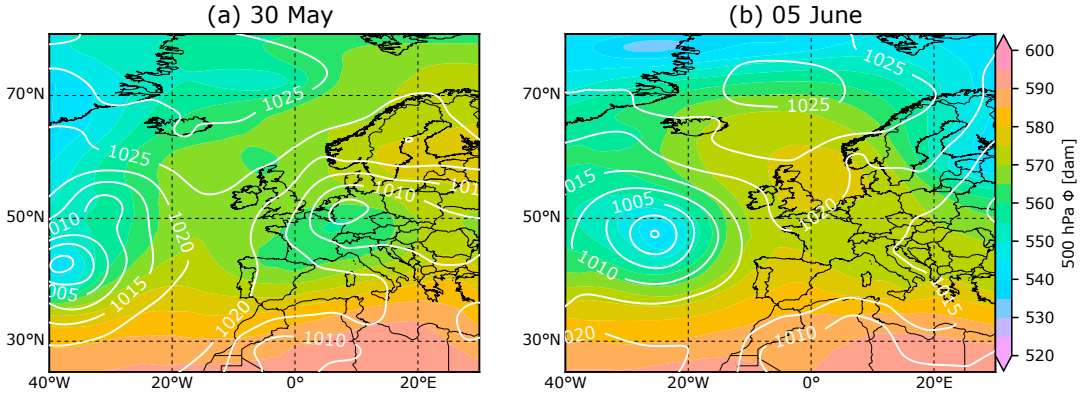


**Figure 11:** Synoptic charts at (a) 00:00 UTC 30 May and (b) 00:00 UTC 5 June 2016. White lines represent mean sea level pressure (hPa). Colors represent 500-hPa geopotential (dam). Figures created from the operational analyses of the ECMWF IFS.

l. 8, 9) and also climate simulations with a km-scale COSMO version exploit this aspect (e.g., Leutwyler *et al.*, 2017). Such biases, however, can pose a severe problem for data assimilation systems (Schraff *et al.*, 2016; Necker *et al.*, 2018), and improvement of deep convection without the loss of temperature skill is desirable. By setting `tur_len` back to its original value, we will examine whether it is possible to achieve realistic convective initiation without this correction.

Initial conditions are taken from the COSMO-KENDA ensemble data assimilation system (Schraff *et al.*, 2016), which is based on a local ensemble transform Kalman filter to directly assimilate conventional observations. Additionally, radar observations are assimilated through latent heat nudging. In our experiments, we only use the deterministic analysis. Using initial conditions provided by a convective-scale data assimilation system reduces the model spin-up compared to downscaled initial conditions. Boundary conditions are provided by global ICON forecasts. Our simulations are started at 00 UTC on each of the simulation days and extend for 24 h.

A simulation without any perturbations, also with `tur_len` = 500 m, is denoted as *Reference*.

## 2.2 SIMULATION PERIOD WITH COSMO

We select a 10-day period from 29 May to 7 June 2016 for our experiments computed with COSMO, which was characterized by heavy precipitation over Germany (Piper *et al.*, 2016). The first half of the period was dominated by an upper-level trough and associated low-pressure systems over Germany and is represented by the 30 May in Fig. 11a. The resulting synoptic lifting combined with South-Easterly advection of moist air caused heavy rainfalls, particularly on 30 May. In the second half of the period, the trough made way for a persistent ridge structure reminiscent of an omega-block. Then the synoptic situation was calm, allowing large instability to build up each day, followed by strong convection. This behavior was particularly dominant on 6 June (Fig. 11b). Due to the intense convective activity and high variability in underlying forcing, this period

has been considered in a number of studies on convection (Rasp *et al.*, 2018b; Baur *et al.*, 2018; Bachmann *et al.*, 2019; Keil *et al.*, 2019) and provides a good testing ground for the developed schemes. For specific purposes, single days from this period are investigated in more detail.

## 2.3 OBSERVATIONS

We compare our simulations to radar-derived precipitation fields. We use the Radar Online Aneichung (RADOLAN) quality-controlled radar observations (EY product) provided by the German Weather Service, which is based on European radar reflectivities and provides radar coverage for most of our domain (Deutscher Wetterdienst, 2018a,b). For all analyses of our simulations, we restrict the geographical region to grid points where radar data is available.

We mostly focus on the systematic impact of the developed schemes on the behavior of the model. The comparison to radar observation only serves as a guideline. Trying to improve the forecast scores would require much longer simulation periods and tuning of several model parameters (e.g., `tur_len`). We anticipate, however, that the systematic impacts of the here considered schemes are independent of the specific model configuration.

## 2.4 VERIFICATION METRICS

As we are mostly interested in convection, we focus on the verification of precipitation. Traditional measures like mean absolute error or the root mean squared error are often not very effective when intermittent, localized precipitation fields are considered. For instance, if precipitation objects are slightly shifted between observations and model simulations, such methods penalize twice: once for regions where precipitation occurs in the observations but not in the simulations, and once where precipitation occurs in the simulations but not in the observations. This problem is often referred to as the double penalty problem. Furthermore, such skill scores easily improve with more precipitation. Alternative measures to quantify the quality of simulated precipitation in comparison to observations, however, exist. Fuzzy verification methods reduce the double penalty by filtering the precipitation fields and thereby allowing for some spatial shifts in the precipitation fields. A widely used fuzzy/neighborhood metric is the so-called fraction skill score, FSS (see Sec. 2.4.1). Other approaches are based on objects. Such methods identify single objects, i.e., clouds or precipitation cells, and compare their characteristics with each other. Such metrics include the Structure component of the Structure-Amplitude-Location (SAL, Sec. 2.4.2) score or simple size and frequency distributions of the identified objects (Sec. 2.4.3). See, e.g., Gilleland *et al.* (2009) for an intercomparison of some methods.

### 2.4.1 *Fraction skill score*

The fraction skill score (FSS) evaluates how two precipitation fields, usually radar observations $O$ and model forecast $F$, agree with each other within a given neighborhood.

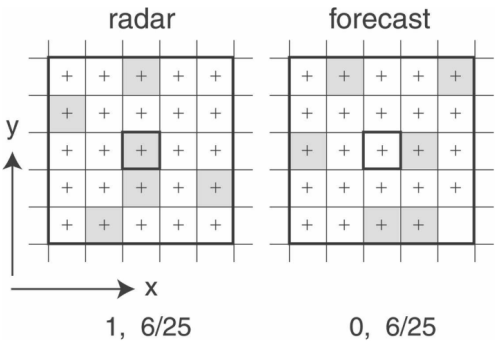


**Figure 12:** Schematic illustration of the neighborhood approach of the FSS. Radar (left) and forecast (right) precipitation is displayed per model grid box in grey shading. The bold big box corresponds to the neighborhood at which the area fraction of precipitation is being evaluated. From Roberts and Lean (2008). ©American Meteorological Society. Used with permission.

First, a binary precipitation field is determined by setting precipitation above/below a certain threshold $q$ to 1/0. Next, the area fraction of this binary precipitation field is computed by accounting for a $n \times n$ box centered around each grid point, as illustrated in Fig 12. These area fractions $\tilde{O}_{(n)}(i,j)$, $\tilde{F}_{(n)}(i,j)$ are then used to compute the ratio of the mean squared errors:

$$MSE_{(n)} = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[ O_{(n)i,j} - F_{(n)i,j} \right]^2, \text{ and}$$

$$MSE_{(n)ref} = \frac{1}{N_x N_y} \left[ \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} O_{(n)i,j}^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} F_{(n)i,j}^2 \right].$$

Then, the FSS is defined as:

$$FSS_{n,q} = 1 - \frac{MSE_{(n)}}{MSE_{(n),ref}}.$$

Values of FSS range from 0 to 1, with a perfect forecast score of 1. With $f_0$ as the area fraction of precipitation in the total domain, Roberts and Lean (2008) suggest that a forecast only contains spatially skillful information if $FSS > 0.5 + \frac{f_0}{2}$. Two parameters, the precipitation threshold $q$ and the neighborhood scale $n$, have to be specified. In general, larger neighborhood scales results in higher FSS values and higher precipitation thresholds at a decrease in FSS. Further details can be found in Roberts and Lean (2008), a comparison with other metrics in Gilleland *et al.* (2009).

### 2.4.2 *The structure component of the SAL-score*

A quantitative measure for the structure of the precipitation field is the S(tructure) component of the SAL score (Wernli *et al.*, 2008). It is based on comparing the normalized area and intensity of precipitation objects between observations (radar) and model output. Precipitation objects are identified using the threshold R*, which is computed based on the suggestion by Wernli *et al.* (2009): $R* = \frac{1}{15} R^{95}$ where $R^{95}$ corresponds to the 95th percentile of precipitation of grid points exceeding $0.1 \frac{mm}{h}$. $R^{95}$ is computed separately for radar and forecast and for each time step. For each identified object $n$

a "scaled volume" $V_n$ is computed as the total precipitation of the object $R_n = \sum_{i,j} R_{i,j}$ normalized by its maximum precipitation rate $R_n^{max}$:

$$V_n = \frac{R_n}{R_n^{max}}.$$

Having computed this scaled volume for all precipitation objects its weighted average can be computed:

$$V(R) = \frac{\sum_{n_1}^{M} R_n V_n}{\sum_{n_1}^{M} V_n}$$

Finally the S-score is computed as the normalized difference in V between model (*mod*) and observations (*obs*):

$$S = \frac{V(R_{mod}) - V(R_{obs})}{0.5\,[V(R_{mod}) - V(R_{obs})]}$$

Negative S values imply that the simulated precipitation cells are too small and peaked compared to the radar observations while $S = 0$ suggests a perfect match in terms of structure. Different threshold choices, such as fixed precipitation amounts, did not change the outcome qualitatively.

The amplitude (A) and location (L) score of the SAL score are also object-based methods. For our purposes, however, we do not find additional value in considering them.

### 2.4.3 *Size and frequency distributions*

It is also possible to simply analyze size or frequency distributions of identified precipitation objects. Objects are then simply identified as a connected agglomeration of grid points with precipitation higher than a specific threshold.

# 3

# REPRESENTING BOUNDARY-LAYER TURBULENCE VARIABILITY

As explained in Sec. 1.3.2, boundary-layer turbulence is strongly coupled to convective initiation. Daytime surface heating stimulates boundary-layer turbulence and thereby warms up the boundary layer via mixing, reduces CIN and preconditions the boundary layer for convective initiation. Stronger turbulent eddies can further push through the convective inhibition, reach their LFC and develop to convective plumes.

The largest turbulent eddies in the boundary layer have characteristic scales corresponding to the boundary-layer height, i.e., around 1 km, and have to be parameterized in NWP models (see Sec. 1.1.3 and 1.2). Originally such boundary-layer parameterizations were developed for grid sizes that are large relative to the turbulent eddies and typically only represent the mean mixing caused by subgrid-scale eddies in a deterministic way. The parameterization predicts the mean effect of a large ensemble of eddies, and the variability associated with the individual eddies is averaged out. For convection-permitting models, however, a km-scale grid box will contain only a small number of the largest eddies, and their variability will have a substantial effect on the grid-box mean (see Sec. 1.2.1 on grey zones). As a result, the variability on the smallest resolved scales in convection-permitting model is under-represented, and difficulties in overcoming convective inhibition arise. This ultimately causes deficits in precipitation forecasts, particularly when synoptic forcing is weak (Keil *et al.*, 2014). These deficits are specifically a lack of convective initiation during the day, a late onset of convection, too frequent grid cell storms, and a shortage of organization (Baldauf *et al.*, 2011; Hanley *et al.*, 2015; Clark *et al.*, 2016; Rasp *et al.*, 2018b).

In this chapter, we aim to improve the representation of convective initiation by boundary-layer turbulence in km-scale models. We use process-level stochastic perturbations, which better represent the subgrid-scale variability and accordingly quantify uncertainty, as explained in Sec. 1.2.2. A corresponding parameterization was originally developed by Kober and Craig (2016), the physically-based stochastic perturbations, i.e., PSP. We revise this scheme to improve its consistency using COSMO. Based on these results, an updated version of the PSP scheme will be defined that has the potential to improve forecast skill for convective precipitation.

The PSP scheme (Kober and Craig, 2016) aims to better couple subgrid-scale turbulence to convective initiation by reintroducing missing subgrid-scale variance to the resolved scales. This is done by adding stochastic perturbations to the model tendencies of temperature, humidity and vertical velocity. The perturbation structure is given by a horizontally correlated random field that is regenerated every 10 minutes, a time span that represents the average lifetime of a boundary-layer eddy. Since such eddies span the entire depth of the convective boundary layer, the perturbations are kept vertically coherent by applying the same random field at all model levels. Most importantly, the amplitudes of the perturbations are scaled by physical information on the subgrid-scale

variance computed in the turbulence parameterization. This scaling links the perturbation strength to the physical processes that cause small-scale turbulence. Furthermore, this enables the scheme to adapt to different atmospheric situations.

The PSP scheme showed promising results in a range of case studies (Kober and Craig, 2016). Under weak synoptic forcing, the perturbations trigger more convective cells and consequently produce more precipitation. This improved the diurnal cycle and amplitude of precipitation. Further, Rasp *et al.* (2018b) have successfully exploited its capability to create different realizations of the convective cloud field for the same synoptic situation by generating different random fields for each ensemble member. The magnitude of the ensemble spread created by different PSP realizations was also found to be comparable to operational ensembles (Keil *et al.*, 2019). And in a convective-scale data assimilation framework, the PSP scheme produced a beneficial impact in ensemble spread and precipitation forecasts (Zeng *et al.*, 2020).

However, the original version of the PSP scheme also leads to undesirable side effects. We investigate these issues and propose modifications to alleviate them. In particular, four topics will be covered:

1. In its current formulation, the evolution of the random field is temporally discontinuous, being regenerated every 10 minutes. To avoid these discontinuities, we introduce a continuously evolving random field using an autoregressive (AR) process.

2. Unrealistically high night-time precipitation has been observed using the PSP scheme (Rasp *et al.*, 2018b). As we will show, this is related to perturbations in the free-troposphere. These do not fit with the rationale of the PSP scheme, which is designed to represent vertically coherent structures in convective boundary layers. Hence to remove these artifacts, we restrict the perturbations to the boundary layer.

3. Furthermore organized convective structures were found to be broken apart by the perturbations. To prevent this, we investigate the use of a mask to switch off the perturbations in regions where precipitation is already occurring.

4. Finally, the vertical velocity perturbations were found to have an insignificant effect compared to the temperature and humidity perturbations (Kober and Craig, 2016). As we will discuss in this paper, this is associated with the generation of acoustic modes, which lead to the rapid dissipation of the imposed vertical velocity perturbations. By introducing balanced, 3D wind perturbations, we obtain lasting and effective wind perturbations.

After evaluating the modifications separately, we propose a revised version of the scheme, which we will call PSP2.

In Section 3.1, the formulation of the PSP scheme is explained in detail, as well as the four modifications mentioned above. In Section 3.2, we describe our strategy for the model. The results are presented in Section 3.3 followed by a summary and discussion in Section 3.4.

## 3.1 CONCEPTION OF THE STOCHASTIC PERTURBATIONS

In this section, we summarize the original perturbation scheme for boundary-layer turbulence (PSP) by Kober and Craig (2016) and then describe our modifications to improve the physical consistency of the scheme.

### 3.1.1 *Physically based stochastic perturbations for subgrid-scale turbulence (PSP)*

The PSP scheme for boundary-layer turbulence was first described in Kober and Craig (2016). Here we present an updated formulation according to Rasp *et al.* (2018b).

The PSP scheme reintroduces the influence of the lost small-scale variability of boundary layer eddies by adding perturbations to the tendencies of the temperature, humidity and vertical velocity field ($T$, $q_v$, $w$) on the smallest effectively resolved scale, $\Delta x_{\text{eff}} = 5\Delta x$ (see e.g., Bierdel *et al.*, 2012). For this we assume a random, horizontal eddy field $\eta(x, y, t)$ with temporal and spatial correlation scales $\tau_{\text{eddy}}$ and $\Delta x_{\text{eff}}$ (see Fig. 13a). The temporal correlation is given by the typical life time of convective eddies, i.e., $\tau_{\text{eddy}} = 10\,min$, and is also theoretically justified by the missing scale separation between resolved and subgrid-scales which necessitates the representation of memory effects (see e.g., Berner *et al.*, 2017, or Sec. 1.2). The perturbation amplitude is chosen to be proportional to the subgrid standard deviation $\sqrt{\overline{\Phi'^2}}$ for $\Phi \in \{T, q_v, w\}$, which is taken directly from the turbulence scheme. The COSMO model uses a 1.5 order closure (Raschendorfer, 2001) based on level 2.5 of Mellor and Yamada (1982), in which the second moments are diagnostically computed based on the turbulence kinetic energy and the vertical gradients of the variables in question. In addition, the scheme adapts to the grid spacing. For uncorrelated subgrid-scale eddies the variability scales with $1/\sqrt{N_{\text{eddy}}}$, where $N_{\text{eddy}} = \frac{\Delta x^2}{l_{\text{eddy}}^2}$ is the number of eddies of scale $l_{eddy} = 1000\,m$ in a grid box (Craig and Cohen, 2006). Dividing by $\tau_{eddy}$ ensures that the accumulated perturbations during this time period scale as $\frac{l_{\text{eddy}}}{\Delta x_{\text{eff}}}\sqrt{\overline{\Phi'^2}}$. The complete mathematical formulation is:

$$\partial_t \Phi|_{\text{PSP}} = \alpha_{\text{tuning}} \eta \frac{1}{\tau_{\text{eddy}}} \frac{l_{\text{eddy}}}{\Delta x_{\text{eff}}} \sqrt{\overline{\Phi'^2}}, \tag{3}$$

where $\alpha_{\text{tuning}}$ is a free parameter that should be of order one, if the length, time and variance scales that appear in the equation have the correct orders of magnitude. A single choice of $\alpha_{\text{tuning}}$ should apply for all weather regimes and model resolutions, if the physical assumptions of the scheme are satisfied. Example fields for the random field $\eta$, the subgrid standard deviation for temperature and the resulting temperature perturbations are shown in Fig. 13.

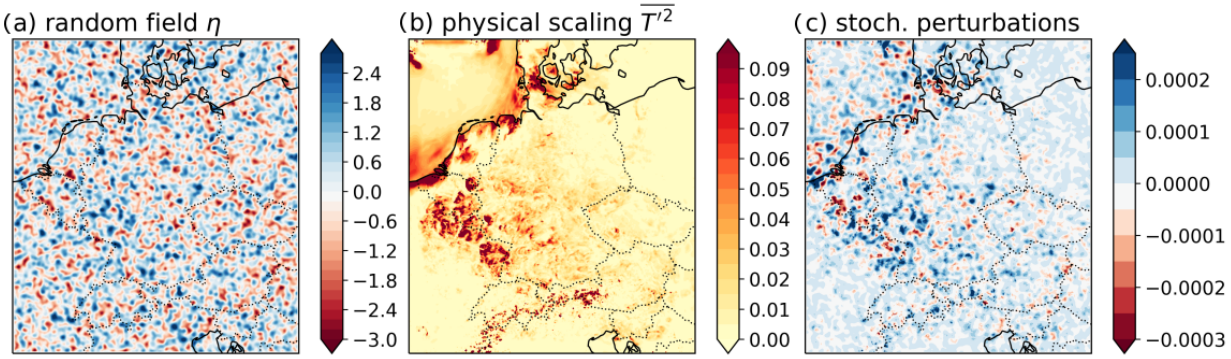


**Figure 13:** Different components for the stochastic perturbations are shown for the 5th lowest model level at 18:00 UTC using the *PSP* simulation (see Table 1). (a) shows the gaussian convolved random field; (b) displays the subgrid variability of temperature (in $[K^2]$); and (c) shows the final temperature perturbations (in $[Ks^{-1}]$).

### 3.1.2 *Modifications to the PSP scheme*

#### 3.1.2.1 *Autoregressive process: PSP-AR*

In the original version of the PSP scheme, a new random field $\eta$ was generated every $\tau_{eddy} = 10\,min$ and was held constant in between, as illustrated in Fig. 14. To ensure a more realistic evolution, we introduce a continuously evolving random field by using an auto-regressive (AR) process:

$$\eta_t = \sigma_\tau \cdot \eta_{t-1} + \sigma_\epsilon \cdot \epsilon_t.$$

The autoregressive parameter $\sigma_\tau$ defines the temporal correlation scale as described below. A random field $\epsilon_t$ is generated at each time step $t$ with same spatial correlation scale as $\eta$, and standard deviation 1 and mean 0. The parameter $\sigma_\epsilon$ scales $\epsilon$ as described below. To ensure a stationary process (i.e., mean and variance remain constant with time) that contains "memory" (i.e., subsequent time steps are positively correlated) we require $0 < |\sigma_\tau| < 1$ (Wilks, 2005). To determine the two scaling parameter $\sigma_\tau$ and $\sigma_\epsilon$ we consider two constraints. First, the variance of the random field $\eta_t$ is required to be 1, which leads to the relationship $\sigma_\epsilon = \sqrt{1 - \sigma_\tau^2}$. Second, the autoregressive parameter $\sigma_\tau$ is defined in terms of the characteristic time scale for decay of the temporal correlation between two fields, $\tau$. We take this correlation time scale $\tau$ as the exponential decay scale of the autocorrelation function $\rho(k) = \sigma_\tau^k = e^{k \ln \sigma_\tau}$, where k corresponds to the lag $t - t_0$. Then, the correlation time scale is linked the autoregressive parameter $\sigma_\tau$ by $\tau = -\frac{1}{\ln \sigma_\tau}$ (see Wilks, 2005, for more details). We set $\tau$ to the time scale at which the random field was modified in the original PSP scheme, i.e., $\tau_{eddy} = 10\,min$. A corresponding time series is shown in Fig. 14. The original PSP scheme with this modification will be denoted as *PSP-AR*.

#### 3.1.2.2 *Adaptive boundary-layer height restriction: PSP-HPBLcut*

It will be shown in section3.3.2 that the original PSP showed unrealistic precipitation at night (Fig.15), which can be related to considerable perturbations in the free tropo-

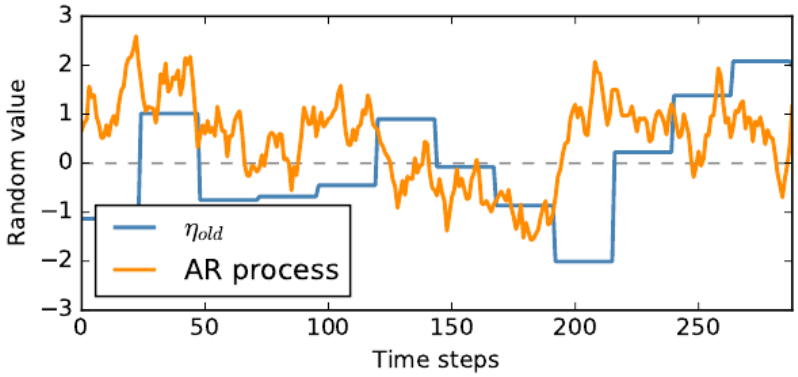


**Figure 14:** Time series of possible random values at single grid points for the original formulation (blue) and with an autoregressive process. One time steps corresponds to 25 seconds.

sphere present also during the night (see Fig. 16). However, the PSP was developed with the convective boundary layer in mind, and many of its assumptions (notably vertically correlated eddies) are not valid for free tropospheric turbulence. So to improve the physical integrity of the scheme and simultaneously reduce unrealistic precipitation, we limit the perturbations to the planetary boundary-layer height $H_{PBL}$. We determine $H_{PBL}$ as the height at which the bulk Richardson number reaches the critical Richardson number as implemented in the COSMO model. This value is 0.33 under stable conditions (Wetzel, 1982) and 0.22 under convective conditions (Vogelezang and Holtslag, 1996). Below $H_{PBL}$, the perturbations are applied as usual. Above, however, the perturbations are decreased linearly to zero at a height of 500 $m$ above $H_{PBL}$. This modification, applied to the PSP-AR version of the scheme, will be termed PSP-HPBLcut.

#### 3.1.2.3 *Masking of already precipitating areas: PSP-mask*

We observed that, in certain situations, the original PSP scheme appeared to break up organized convection (see Section 3.3.3). We hypothesize that this is partly caused by the perturbations disturbing already existing updrafts. As a simple *ad hoc* fix, we test turning off the perturbations in grid columns where the vertically integrated precipitating hydrometeor content exceeds $10^{-3}\,kg\,m^{-2}$. This adjustment to the PSP-AR version will be labeled as PSP-mask.

#### 3.1.2.4 *3D non-divergent perturbations: PSP-3D*

Kober and Craig (2016) observed that the impact of the original PSP scheme was caused by the temperature and humidity perturbations while the vertical velocity perturbations had essentially no effect. Further analysis shows that the $w$ perturbations decay rapidly in a few time steps. This can be related to pressure perturbations: Chagnon and Bannon (2005), for example, show that an injection of vertical momentum efficiently induces high-frequency acoustic waves causing a rapid removal of the imposed vertical updraft. Following their approach, we extend the scheme to add perturbations to the horizontal velocity field that approximate three-dimensional non-divergence. The goal is to suppress the excitation of spurious acoustic waves so that the vertical velocity perturbations persist (see e.g., Fiedler, 2002; Chagnon and Bannon, 2005; Edson and Bannon, 2008).

Non-divergence is given by:

$$\nabla \cdot \vec{v} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0.$$

We neglect modifications due to the terrain following vertical coordinate system of the COSMO-model and the metrical terms due to the spherical earth as well as compressibility. Neglecting the distortion to the grid due to these effects greatly simplifies the computation, but implies that the resulting wind perturbations will not be exactly non-divergent, and some perturbation energy will be lost to acoustic waves. However, this lost perturbation energy will be small in comparison to the lost perturbation energy of the original PSP scheme, where the wind perturbation was entirely divergent, and almost all the energy was rapidly dissipated. The COSMO-model employs a rotated spherical coordinate system with earth radius $R$, rotated longitude $\lambda$ and latitude $\phi$, so that $\partial x = R \cos\phi \partial\lambda$ and $\partial y = R\partial\phi$. Due to the rotated coordinate system and the comparably small domain, we neglect the trigonometric terms for simplicity.

This results in the following approximate non-divergence criterion for the PSP setup:

$$-R \cdot \frac{\partial w}{\partial z} = \left[\frac{\partial u}{\partial \lambda} + \frac{\partial v}{\partial \phi}\right] \tag{4}$$

We define the horizontal perturbations by the velocity potential $\chi$: $\vec{v}_h = \nabla\chi$. Consequently, the horizontal perturbations are non-rotational, which is in accordance with the incentive of the perturbations.

Computing the divergence of $\vec{v}_h$ and substituting in equation 4 we obtain the following Poisson equation:

$$-R^2 \cdot \frac{\partial w}{\partial z} = \left[\frac{\partial^2 \chi}{\partial \lambda^2} + \frac{\partial^2 \chi}{\partial \phi^2}\right] \tag{5}$$

For a vertical velocity perturbation (increment) given by eq. 3 we derive $u$ and $v$ perturbations by solving (5) for the corresponding perturbation of $\chi$. A finite-difference approximation is obtained by replacing the Laplace operator with a five-point-stencil Laplace matrix.

To avoid large vertical velocity gradients $\frac{\partial w}{\partial z}$ near the lower boundary, we linearly increase $w$ perturbations from 0 at the surface to the original value at $z_0$. We chose $z_0 = 500\,m$ to include a sufficient number of vertical model levels while ensuring that the full perturbation amplitude is active at the height of the daytime inversion, where we expect the perturbations to be most effective.

Applying these balanced wind perturbations to PSP-AR will be denoted as *PSP-3D*.

### 3.2 STRATEGY FOR EVALUATING THE IMPACT OF THE PSP MODIFICATIONS

To evaluate the impact of the PSP variants, we compute several simulations with different setups of the COSMO model. We investigate each modification of PSP individually on one day with weak synoptic forcing of convection, 6 June 2016. For comparison, we use 30 May, 2016 as an example of a strong forcing day with synoptically driven, coherent precipitation. Based on these results, we decide on a suitable configuration for a

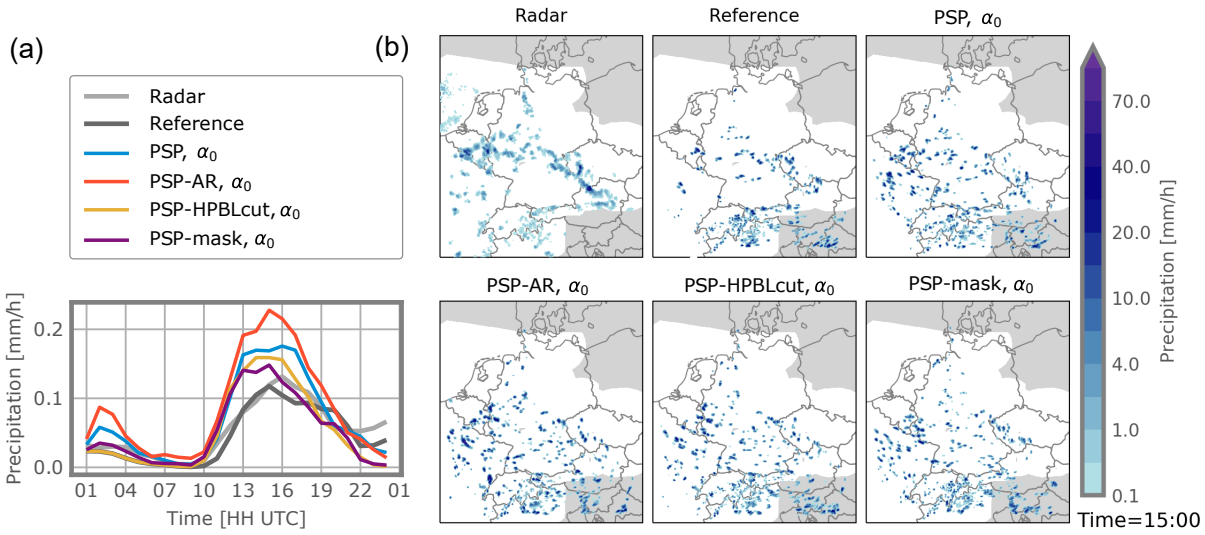


**Figure 15:** (a) Hourly accumulated domain-mean precipitation for 6 June 2016. The domain-mean is computed only for the region of available radar data. (b) Snapshots of hourly precipitation accumulation at 1500 UTC on 6 June 2016. Grey shading indicates missing radar data. The different experiments are listed in Table 1 (The perturbed simulations shown were computed using $\alpha_0 = 7.2$).

revised PSP version, PSP2, and compute 10 subsequent days with total simulation time of 24h to confirm our findings. See Table 1 for an overview of the experiments. Details on COSMO, the selected days, the used radar observations and verification metrics can be found in Chapter 2.

**Table 1:** Overview of the considered experiments and corresponding model settings.

| | Pert. variables | | | | | |
|---|---|---|---|---|---|---|
| Exp. name | T, $q_v$ | w | u-v | rand. field | $\alpha_{\text{tuning}}$ | Description |
| *PSP* | yes | yes | no | const. for $\tau$ | $\alpha_0 = 7.2$ | As in Kober and Craig (2016) |
| *PSP-AR* | yes | yes | no | AR | $\alpha_0 = 7.2$, $\alpha_1 = 2$ | Auto-regressive random field |
| *PSP-HPBLcut* | yes | yes | no | AR | $\alpha_0 = 7.2$ | Pert. go to zero above BL height |
| *PSP-mask* | yes | yes | no | AR | $\alpha_0 = 7.2$ | No pert. in precipitating columns |
| *PSP-AR* $Tq_v$ | yes | no | no | AR | $\alpha_1 = 2$ | Only pert. to thermodynamic variables |
| *PSP-3D w* | no | yes | yes | AR | $\alpha_1 = 2$ | Only 3D non-divergent wind pert. |
| *PSP-3D all* | yes | yes | yes | AR | $\alpha_1 = 2$ | Pert. to all fields |
| *PSP2* | yes | yes | yes | AR | $\alpha_{final} = 1.5$ | Comb. modifications (3D all + HPBLcut) |

## 3.3 RESULTS

### 3.3.1 *Autoregressive Process in PSP-AR*

To test the influence of the AR random field, we focus on 6 June, a weakly forced day on which the perturbations are expected to have a large influence on the precipita-

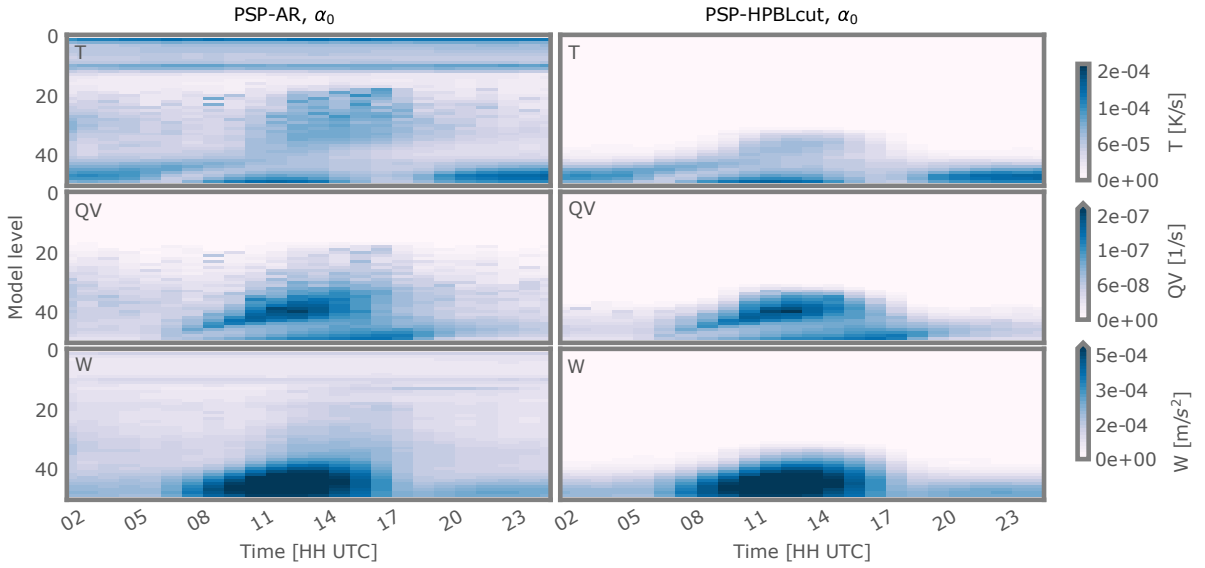


**Figure 16:** Standard deviation of tendency perturbations over the horizontal domain for each vertical model level and hourly timesteps: temperature (top row), specific humidity (middle row) and vertical velocity (bottom row). The left/right columns show the perturbations for the PSP-AR/PSP-HPBLcut simulations.

tion. The original PSP scheme (*PSP*), in which the random field is regenerated every 10 minutes, leads to an increase and earlier onset of precipitation (Fig. 15a) compared to the unperturbed reference simulation. We note that the diurnal cycle and amplitude of precipitation from radar observations (Fig. 15a) are already well matched by the *Reference* simulation. The *PSP* simulations consequently produce too much precipitation. This behavior is, however, not necessarily typical of longer periods (see Section 3.3.5). As mentioned above, due to the tuning problem of NWP, we concentrate on the systematic impacts of the PSP schemes, not their performance compared to observations. Switching to a continuously evolving random field (*PSP-AR*) further increases the precipitation amount. Example precipitation fields in Fig. 15b confirm that this increase in precipitation is a consequence of more convective cells and, thus, more triggering.

We explain this behavior by the different meaning of $\tau_{eddy}$ in the *PSP-AR* setup: For the AR process, $\tau$ represents the average decay scale, while for single grid points and time ranges longer correlation times are possible. These might be more effective in triggering precipitation. The *PSP-AR* perturbation can be scaled by tuning $\alpha_{\text{tuning}}$, which we will do after also considering the other modifications. Note also that these results are robust for different random seeds (not shown). In the remainder of this paper, we will always use the AR random field, unless otherwise specified.

### 3.3.2 *Adaptive boundary-layer height restriction in PSP-HPBLcut*

Next, we investigate the impact of constraining the perturbations to the height of the planetary boundary layer (*PSP-HPBLcut*). Fig. 16 shows the horizontal standard deviations of the perturbations. Without limiting the perturbation height (*PSP-AR*) the

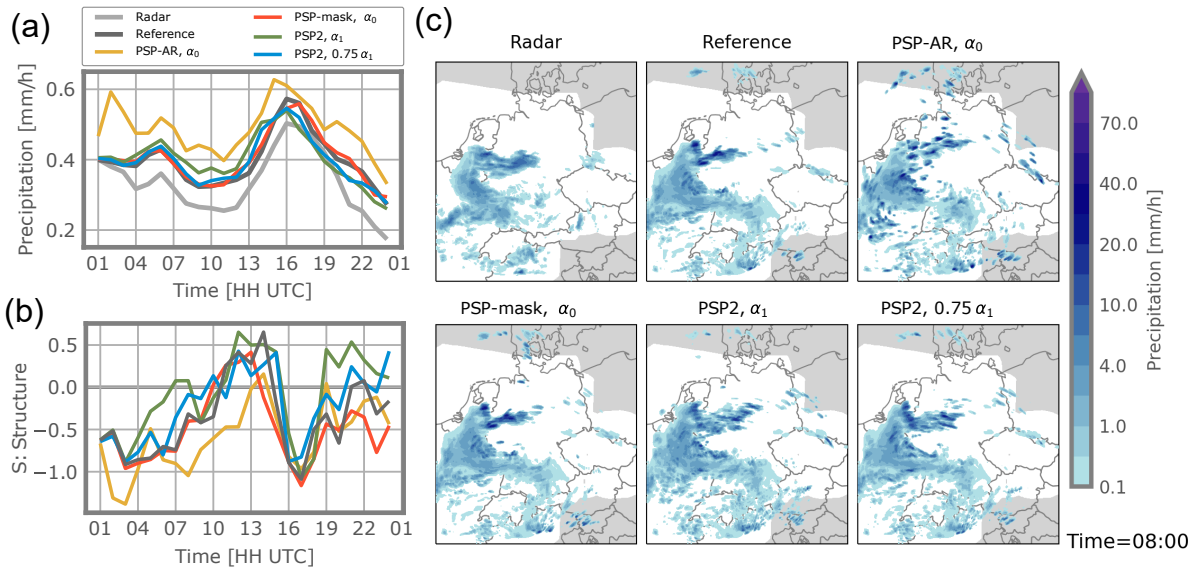


**Figure 17:** Illustrating the impact of precipitation mask: time evolution for the 30 May 2016 for (a) domain averaged precipitation and (b) structure component of SAL. (c) Example precipitation fields for 30 May 2016, 08 UTC.

perturbations are largest in the boundary layer, but significant perturbations in the mid-troposphere are present as well. As discussed above, this is conceptually inconsistent with the rationale of the PSP scheme. Limiting the height in *HPBLcut* removes this inconsistency. This modification leads to a decrease in the total precipitation amount and, most importantly, removes the precipitation peak at night (Fig. 15a).

### 3.3.3 *Masking of already precipitating areas in PSP-mask*

In the original and the PSP-AR scheme, we observed a breaking up of larger-scale precipitation structures. An example for this can be seen on 30 May (Fig. 17c), a strongly forced convective day. A coherent, synoptically-driven precipitation structure is present over Western Germany and Belgium, which is well reproduced by the reference simulation. Adding the *PSP-AR* perturbations breaks up this structure into many smaller convective cells. Turning off the perturbations in precipitating columns, *PSP-mask* corrects this problem but also reduces the impact on weak forcing days (see Fig. 15). A too-small threshold for precipitating hydrometeors can be responsible for switching off the perturbations before deep convection has developed.

A quantitative measure for the structure of the precipitation field is the S(tructure) component of the SAL score (Wernli *et al.*, 2008) displayed in Fig. 17 b. Negative S values imply that the simulated precipitation cells are too small and peaked compared to the radar observations, while $S = 0$ suggests a perfect match in terms of structure (see 2.4.2 for details). During the morning hours on 30 June, when the synoptic structure is most salient, the *PSP-AR* simulations show a clear reduction of the S score, indicating smaller cells. The precipitation mask reverts the structure back to its original value.

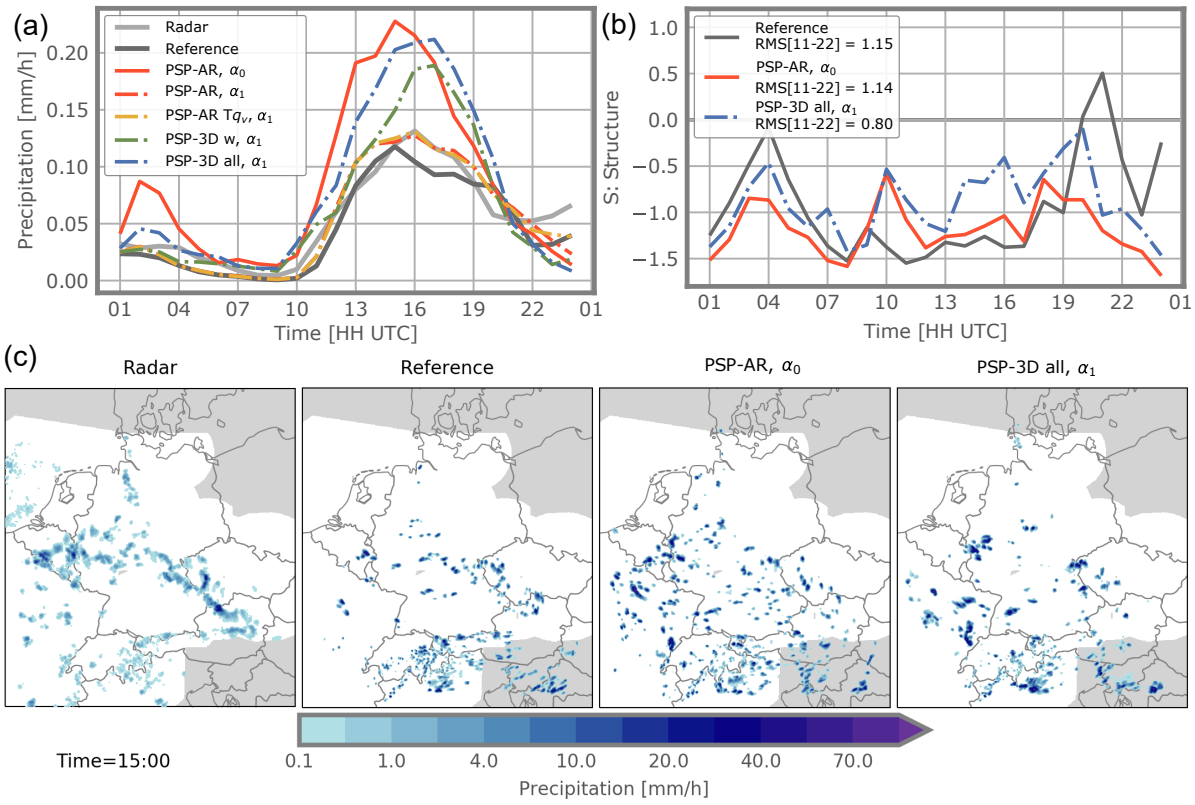


**Figure 18:** (a) As Figure 15a but considering different PSP setups for analysing the 3D modification. Note that $\alpha_1 = 0.28\alpha_0$. (b) Structure component $S$ of SAL for selected simulations as indicated by colors and line style. The RMS between 11 and 22UTC, when most precipitation occurs, is given in the legend as well. Values close to zero indicate similar structure as radar observations. Positive/negative $S$ implies that the simulated precipitation cells are too small/widespread when compared to radar observations. (c) As Fig. 15b but for different PSP setups.

The precipitation mask solves the problem of the original scheme, where organized convection was broken up, but as noted above, it also reduces the triggering effect in weak forcing situations.

### 3.3.4 *3D non-divergent perturbations in PSP-3D*

As a last modification to the PSP scheme, we investigate the impact of adding non-divergent wind perturbations. For these experiments, we reduce the perturbation amplitude $\alpha_{\text{tuning}}$ to $\alpha_1 = 2$ because larger horizontal wind perturbations resulted in numerical instabilities. The domain averaged precipitation time series are displayed in Fig. 18a. As expected, the precipitation increase is lower for the new perturbation amplitude $\alpha_1$ compared to the original perturbation amplitude $\alpha_0$ (see Tab. 1). We first investigate the impact of the original $w$ perturbations without horizontal wind perturbations by comparing *PSP-AR*, $\alpha_1$ to *PSP-AR* $Tq_v$, $\alpha_1$, where only temperature and humidity are perturbed. Consistent with the results of Kober and Craig (2016), $w$ has a negligible effect in this experiment.

Adding the $u$ and $v$ components to enforce non-divergent wind perturbations results in a marked increase in precipitation even without the $T$ and $q_v$ perturbations (*PSP-3D w*, $\alpha_1$). Including perturbations to $T$ and $q_v$ again (*PSP-3D all*, $\alpha_1$) further increases the effect . This suggests that the effect of non-divergent wind perturbations is now of the same order of magnitude as the T and q perturbations. We emphasize that the relative amplitudes of the imposed wind and thermodynamic perturbations are determined from the subgrid variances derived from the turbulence scheme.

We further investigate the impact of the non-divergent wind perturbations on the precipitation structure (Fig. 18b, c). *PSP-3D all*, $\alpha_1$ has fewer but larger convective cells compared to *PSP*, $\alpha_0$ at a comparable total precipitation amount. Again we look at the S score to quantify the convective structure. In general, the model simulations show negative S values. This confirms the visual impression that the cells are too small and intense and lack the stratiform precipitation regions in the observations. During the daytime the PSP-AR simulation has slightly improved scores compared to the unperturbed reference simulations. The non-divergent wind perturbations result in a further improvement, particularly during the time of maximum convection in the afternoon. Hence, balanced wind perturbations lead to larger cell structures.

In general, the non-divergent wind perturbations are very effective in triggering convection and also appear to induce mesoscale circulations that cause larger precipitation cells.

### 3.3.5 *Combining the modifications for PSP2*

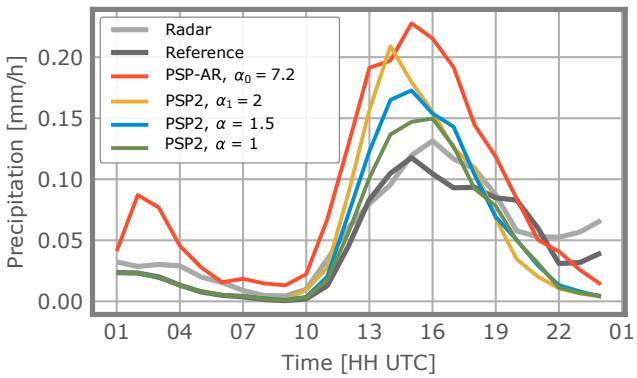


**Figure 19:** Domain averaged, hourly accumulated precipitation for 6 June considering different amplitudes for the new PSP2 setup.

Finally, we choose the most suitable combination of modifications to PSP for further investigation. This includes the AR random field, HPBLcut and the non-divergent wind perturbations and will be denoted as PSP2. Since the 3D wind perturbations already improve the precipitation structure (see Fig. 17c, PSP2), we do not include the precipitation mask, which has the undesirable side effect of reducing convective initiation. Because of the many changes to the original scheme, we start by determining a reasonable perturbation amplitude $\alpha_{\text{tuning}}$. Corresponding domain averaged precipitation time series are displayed in Fig. 19 for $\alpha_{\text{tuning}} = 1, 1.5, 2$. We select $\alpha_{final} = 1.5$ which produces perturbation amplitudes comparable to the original PSP (see blue line in Fig. 15) and use this value henceforth.

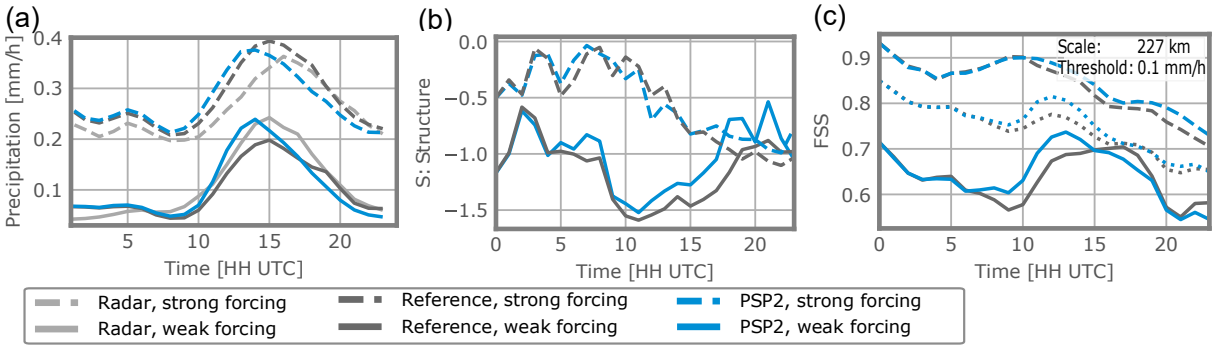


**Figure 20:** (a) Daily cycle of hourly accumulated, domain averaged precipitation separated into a weakly and a strongly forced period. (b) Daily cycle of S(tructure) component of SAL for the weakly forced period and (c) FSS for the whole period (dotted), strong (dashed) and weak (solid) forcing. The FSS was computed using a threshold of $0.1\,mmh^{-1}$ and a scale of $227\,km$.

We now simulate the entire high-impact weather period. For the analysis, we divide it into a strong forcing period (29 May – 2 June) and a weak forcing period (3 – 7 June). For both synoptic situations, the modified PSP2 scheme produces an earlier peak of precipitation (Fig. 20a). This effect is desirable, since it addresses the known problem of late initiation of convection in km-scale models. However, it should be noted that for this particular model and this weather period, the change does not result in an overall improvement in the precipitation forecast in comparison to radar observations.

For the weak forcing days, the precipitation amplitude is also increased, while for the strong forcing days, the amplitude is decreased slightly. This reduction could be caused by a reduction of CAPE in the evening due to earlier triggering or the 3D wind perturbations.

We again look at the S component of the SAL score to investigate the impact on precipitation structure (Fig. 20b). For the weak forcing period, the PSP2 scheme shows a shift towards more coherent precipitation. This is a change in the right direction, but the convective cells are still too small-scale compared to observations.

Finally, in Fig. 20c, we look at a common metric for assessing precipitation forecast skill, the fraction skill score (FSS). Displayed is the FSS for a scale of $227\,km$ with a precipitation threshold of $0.1\,mmh^{-1}$. Results have also been computed for other scales and thresholds (not shown). As expected, the skill decreases for smaller scales and higher thresholds; however, the relative performance of the reference and PSP2 forecasts is similar for all values. In particular, the PSP2 perturbations lead to a slight improvement in skill for the strong forcing days, particularly during the time of maximum convective activity. For the weak forcing days, there is a skill increase during the time of convective initiation but a small decrease of skill towards the evening hours. This decrease is likely connected with the too early decrease in precipitation caused by the perturbations, as discussed above. To evaluate the significance of these differences, longer periods need to be investigated.

To summarize, the new version, PSP2, includes the AR random field, the restriction of the perturbations to the boundary layer, and the balanced wind perturbations. As a preliminary estimate, we re-tune the perturbation amplitude $\alpha_{\text{tuning}}$ to 1.5 to match

the precipitation amplitude of the original PSP scheme. For operational use, this choice would have to be revisited to produce the best overall precipitation forecast over a long test period. Although a clear overall improvement in forecast skill scores was not demonstrated in the preliminary evaluation here, the specific impacts of the changes to the scheme were as expected. In particular, PSP2 maintains the desired effect of the original PSP scheme, namely an earlier peak and increased amplitude of convection precipitation in weak forcing situations, while several drawbacks have been eliminated. In addition, it improves the cell structure and the spatial forecast skill compared to the unperturbed simulation.

## 3.4 SUMMARY AND DISCUSSION

Four concerns of the original PSP scheme were addressed and corrected by modifying different aspects of the original PSP scheme:

1. The autoregressive process in PSP-AR generates a physically more realistic, continuous evolution of the random field and hence refines the concept of the scheme. The performance is qualitatively comparable to the original PSP.

2. In PSP-HPBLcut, we constrain vertically correlated perturbations to the boundary layer in accordance with the vertical extent of boundary-layer eddies. This removes perturbations in the free troposphere that resulted in spurious night time precipitation.

3. By excluding perturbations in precipitating regions in PSP-mask, we are able to reestablish organized convective structures. However, we decided not to include the PSP-mask adjustments in the revised PSP2 version for two reasons: first, the other modifications improved the break-up of organized structures already; second, this ad hoc fix has the undesirable side effect of reducing convective initiation overall.

4. In PSP-3D, we included 3D non-divergent wind perturbations to prevent the rapid dissipation of vertical velocity perturbations. Interestingly, these consistent wind perturbations are at least as effective as the $T$ and $q_v$ perturbations in triggering convection. The relative importance of buoyant and mechanical lifting in initiating convective updrafts in nature is still unknown. Both processes were found by Torri *et al.* (2015) to play a role in secondary initiation by cold pools. It also appears that the cell structures in the simulations with non-divergent wind perturbations were more comparable to radar observations. The relationship between triggering processes and the resulting convective structures is an important topic for future research.

We combined the first, second and fourth modification to define the PSP2 scheme and selected a tuning parameter of $\alpha_{\text{tuning}} = 1.5$ to match the precipitation amplitude of the original PSP. The optimal value of the tuning parameter is of order one, which is an encouraging sign that our physical reasoning is appropriate. By evaluating PSP2 on a longer time period, we find its behavior well in accordance with the rationale of

the original scheme: when synoptic forcing is weak, we observe a small shift in the diurnal cycle and an increase in total precipitation amplitude; this could counteract long-standing biases in current NWP systems (Baldauf *et al.*, 2011; Clark *et al.*, 2016). During strongly forced episodes or at night, the scheme has only a small impact.

In addition to the systematic impact of the PSP scheme on convective initiation, studies have also demonstrated its ability to increase ensemble spread (Rasp *et al.*, 2018b; Keil *et al.*, 2019). This has the potential to help reduce the under-dispersion in many current convective-scale ensemble forecasting and data assimilation systems (see, e.g., Berner *et al.*, 2017; Ollinaho *et al.*, 2017). Direct radar assimilation might benefit especially from such a perturbation scheme where sufficient spread in the first guess ensemble is particularly necessary. One remaining, undesired effect of the PSP2 scheme is a decrease of precipitation towards the evening. We note that this has also been identified for the original PSP by Rasp *et al.* (2018b), who related this to a lack of convective organization. The organization of convection is caused by processes other than the ones addressed with the PSP2 scheme, most notably cold pools. A better representation of these features and their ability to trigger new convection might be necessary to capture organized evening convection. This will be addressed in in Chapter 5 and 6.

Overall, we conclude the PSP revisions by recommending PSP2 instead of the original PSP: PSP2 contains comparable advantages as the PSP while being physically more consistent.

# 4

# REPRESENTING LIFTING BY SUBGRID-SCALE OROGRAPHY

In this chapter, we address the influence of subgrid-scale orography (SSO) on triggering convection as a complementary process to turbulence. As described in the introduction (Sec. 1.3.3), the importance of orography on convective initiation is unquestioned. Many different mechanisms contribute to orographic convection, including thermally driven slope winds or mechanical lifting, but specific details are less understood. In addition, the relevance of small scale orography, as a proxy for subgrid-scale orography in km-scale models, has not been universally quantified (Tucker and Crook, 2005; Kirshbaum *et al.*, 2007b,a; Langhans *et al.*, 2011; Schneider *et al.*, 2018). Nonetheless, Kirshbaum *et al.* (2018) emphasize the need to account for SSO in convection parameterizations and for developing scale-aware parameterizations in the mountain grey zone. This is further motivated, as SSO can potentially act as a source of predictability (Bachmann *et al.*, 2019, 2020). While most NWP models include some form of parameterization for SSO, vertical motion responsible for triggering convection by SSO, however, is not addressed in these parameterizations (see Sec. 1.3.3).

Consequently, we propose stochastic perturbations to account for convective initiation by subgrid-scale orography. We will name it the SSOSP scheme. We focus on the effects of mechanical lifting and the resulting gravity waves. This process is chosen as the most direct effect of orography among those mentioned in Sec. 1.3.3: the flow over a mountain causes an upward displacement of the air parcels, i.e., a vertical velocity perturbation, which may decay with height, or propagate as gravity wave away from the source, depending on the stratification of the atmosphere. The upward displacement may cause the parcels to reach their level of free convection (LFC) and thereby initiate deep convective updrafts. See Fig. 21 for an illustration.

The new scheme closely follows the formulation of the PSP scheme: vertical velocity tendencies are randomly perturbed with an amplitude that scales with theoretical gravity waves excited by SSO, and horizontal wind fields are accordingly perturbed to yield 3D non-divergence, as illustrated in Fig. 21. We present a considerably revised scheme of one already developed by Brundke (2015) as part of his Master thesis.

Investigating the impact of this scheme will further allow us to estimate the importance of SSO for convective initiation, thereby contributing to the discussion mentioned in the introduction.

We will first introduce the concept of the SSOSP scheme in Sec. 4.1. Then, in Sec. 4.2, we will explain the model simulations to evaluate the scheme. The corresponding results are presented in Sec. 4.3, followed by a summary and discussion in Sec. 4.4.

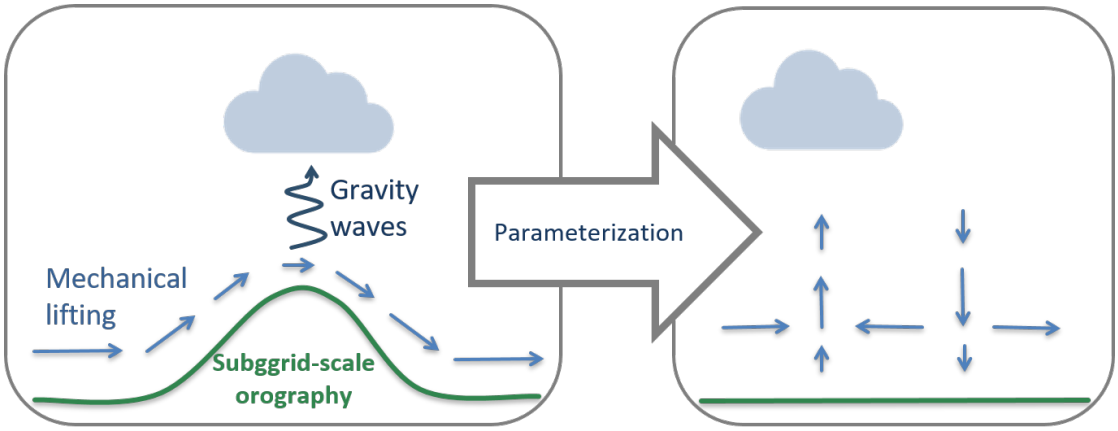


**Figure 21:** Schematic illustration of the SSOSP scheme. The left box illustrates the physical process of mechanical lifting by subgrid-scale orography. The right box shows, how this process is parameterized with SSOSP.

## 4.1 FORMULATION OF THE PARAMETERIZATION

### 4.1.1 *The SSOSP scheme*

Inspired by the PSP scheme, we introduce stochastic vertical velocity perturbations that scale with the amplitude of the SSO lifting and the induced gravity waves. The horizontal wind is perturbed to approximate non-divergence, as described in Section 3.1.2.4. Since the errors in the estimation of the non-divergent wind perturbations are likely to be largest over strong orography where the grid is most distorted, there will be an additional source of noise in these regions. However, we were unable to identify any obvious effects in the simulation results.

The tendency of vertical velocity $w$ is perturbed according to

$$\left.\frac{\partial w}{\partial t}\right|_{sso} = \frac{\alpha_{\text{tuning}}}{\tau} \cdot \eta \cdot w'. \tag{6}$$

$\eta(\tau_\eta, \sigma)$ denotes a horizontally correlated AR random field as described in Sec. 3.1.2.1. Instead of $\tau = \tau_{eddy}$ we assume that slower processes such as the synoptic wind field and the diurnal cycle dominate the evolution of orographically induced flow structures. Craig and Selz (2018) for example show that the time scales of gravity waves (even on scales of 1 km) can range from minutes to days. Since no single time scale stands out, we test several values, $\tau = 30\,min, 2\,h, 5\,h$. As for the PSP scheme, $\alpha_{\text{tuning}}$ is a tuning parameter; ideally, $\mathcal{O}(1)$ if the physical scaling is appropriate.

The variable $w'$ represents the physical scaling of the perturbations. It depends on the amplitude $w_0$ (at given height) and vertical wavenumber $m$ of a gravity wave induced by the SSO. This includes both propagating and decaying gravity waves and also represents the effect of mechanical lifting. Both $w_0$ and $m$ are considered in terms of forced gravity waves as in Gill (1982) and approximated with available SSO information. Following Baines and Palmer (1990), the subgrid-scale orography for one grid box of the COSMO model is represented by four parameters, namely $\theta_{sso}$, $\gamma_{sso}$, $\sigma_{sso}$ and

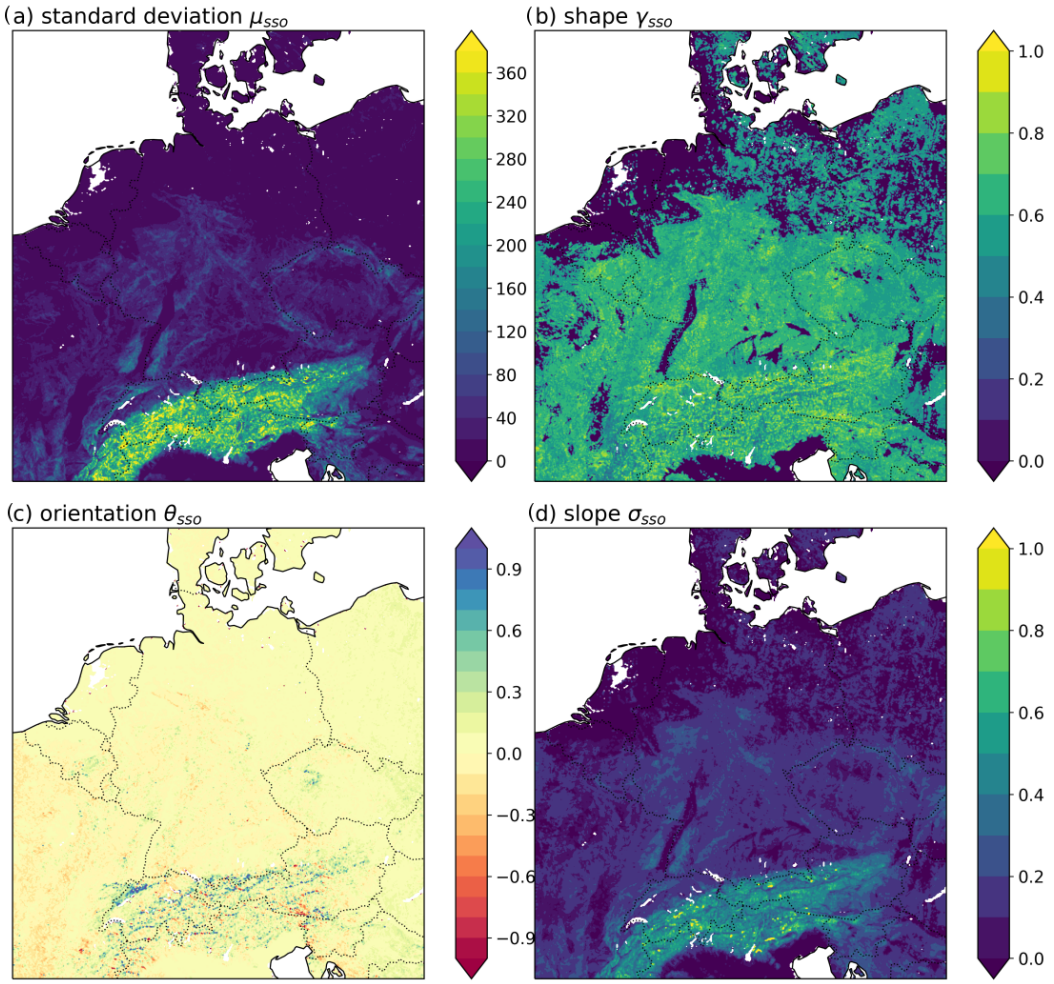


**Figure 22:** Four fields describing the subgrid-scale orography information. These are the standard deviation $\mu_{sso}$ [m], the shape $\gamma sso$, the orientation $\theta_{sso}$ and the slope $\sigma sso$, and $\mu_{sso}$. The data is derived from 30 m ASTER elevation data (Schättler *et al.*, 2016).

$\mu_{sso}$. They describe the orientation, shape, slope and standard deviation of the height of the orography, respectively, and are displayed in Fig. 22. The specific formulation for $w_0$ and $m$ in terms of horizontal wind speed, Brunt-Väisälä frequency $N$ and SSO information is presented in the next two sections.

To avoid numerical instabilities near the surface, we decrease the perturbations linearly to zero between $z_{max} = 500\,m$ and the surface. We further constrain the perturbations to the planetary boundary layer, where we assume triggering processes to be most relevant: we impose an exponential decay of the perturbations at $z_{max}$ that scales

to $\gamma = \frac{1}{2km}$ for non-damped waves and to $max(\gamma_0, \sqrt{-m})$ for damped waves. Mathematically, this can be expressed as:

$$w' = w_0 \cdot f(z),$$

$$\text{where } f(z) = \begin{cases} e^{-\gamma \cdot (z - z_{max})}, & \text{if } z \geq z_{max} \\ \frac{1}{z_{max}} \cdot z, & \text{if } z < z_{max} \end{cases}$$

$$\text{with } \gamma = \begin{cases} \gamma_0 & \omega^2 \leq N^2 \text{ (nondamped)} \\ max(\gamma_0, \sqrt{-m}) & \omega^2 \geq N^2 \text{ (damped).} \end{cases}$$

Details on $w_0$ and $m$ are given in the following section. First, the formulation for orographically induced gravity waves is given. Then we approximate the gravity wave amplitude within the framework of the subgrid-scale orography for the parametrization.

### 4.1.2 *Physical scaling based on orographic gravity waves*

Orography acts as a boundary and creates internal gravity waves, which can be evanescent or propagating. The orographic height can be idealized as a two-dimensional sine wave with amplitude $h_0$ and horizontal wavenumbers $k$ and $l$ (or a composition of many such sine waves). Using a general exponential ansatz for vertical velocity $w$ as solution to the Taylor-Goldstein equation and $\omega$ as oscillation frequency, the following two cases can be distinguished (Gill, 1982): for the propagating case, i.e., $\omega^2 \leq N^2$, the solution for $w$ reads:

$$w = w_0 \cdot \cos(kx + ly + mz - \omega t); \tag{7}$$

when $\omega^2 > N^2$, the amplitude of the vertical displacement decays with height:

$$w = w_0 \cdot e^{-\gamma z} \cdot \cos(kx + ly - \omega t), \tag{8}$$

with $\gamma^2 = -m^2$.

For both cases, $\omega = ku + lv$ and $w_0$ and $m$ are given as:

$$w_0 = (ku + lv)h_0 \tag{9}$$

$$m = (k^2 + l^2)\left(\frac{N^2}{(ku + lv)^2} - 1\right), \tag{10}$$

where $u, v$ denote x- and y-components of the horizontal wind. Further details can be found in Gill (1982).

For the parameterization, we are not interested in representing the complete wave structure. Instead, we aim at representing the effect of SSO-induced gravity waves on the resolved scales in a stochastic manner by scaling the random perturbations accordingly. $w'$ of eq. 6 is hence linked to the amplitude of gravity waves triggered by SSO (see eqs. 7 and 8). To do so, the amplitude of gravity waves triggered by orography (eqs. 7 and 8) is approximated and formulated with the available SSO variables. The cosine

terms in (7) and (8) are consequently neglected. $w_0$ and $m$ are formulated in terms of SSO information. We specify the velocities $u, v$ as $u'$, $v'$ in the rotated system of the SSO:

$$u' = u \cdot \cos\theta_{sso} + v \cdot \sin\theta_{sso}$$
$$v' = v \cdot \cos\theta_{sso} - u \cdot \sin\theta_{sso}.$$

The horizontal wave numbers $k$ and $l$ (in the rotated system of the SSO) are obtained by using the semi-minor and semi-major axes $a$, $b$ for the elliptical mountains that result when the orography is represented by sine waves:

$$k = \frac{\pi}{2a} \approx \frac{\pi}{2}\frac{\sigma_{sso}}{\mu_{sso}}$$
$$l = \frac{\pi}{2b} \approx \frac{\pi}{2}\frac{\sigma_{sso}\gamma_{sso}}{\mu_{sso}}$$

(Lott and Miller, 1997). We set $h_0$ to be $\mu_{sso}$, implying that higher mountains will result in larger vertical displacements.

Using these approximations, we can formulate the vertical wavenumber $m$ and wave amplitude $w_0$, given by (10) and (9), in the following way:

$$m^2 = (1+\gamma_{sso}^2) \cdot \left( \frac{N^2}{(u' + v'\gamma_{sso})^2} - \frac{\pi^2\sigma_{sso}^2}{4\mu_{sso}^2} \right),$$
$$w_0 = \frac{\pi}{2} \cdot \sigma_{sso} \cdot (u' + v'\gamma_{sso}).$$

$w_0$ and $m$ can now be used as physical scaling for the stochastic parameterization.

## 4.2 MODEL SIMULATIONS, OBSERVATIONS AND SIMULATION PERIOD

As for the PSP scheme, we evaluate the impact of the SSOSP scheme by computing several simulations with different setups using the COSMO model as described in Chapter 2. We only consider 6 June 2016 as a day with weak synoptic forcing (see Chapter 2). First, we test several settings of the SSOSP scheme, namely the impact of the correlation time scale $\tau$ and the perturbation amplitude $\alpha_{ref}$, where $\alpha_{ref} = \frac{\alpha_{\text{tuning}} \cdot 2h}{\tau}$. This re-scales the final perturbation amplitude for different values of $\tau$ with respect to a reference timescale $\tau_{ref} = 2\,h$. We identify the most appropriate setting and compute an ensemble simulation with 10 members, where different random seeds of the random field $\eta$ are used.

To see the effects of the orography more clearly, we analyze orographic and flat grid points separately by defining orographic/flat grid points as gridpoints where the SSO standard deviation $\mu_{sso}$ is larger/lower than 20 m. This threshold approximately coincides with the median of SSO-std height. Separating Southern and Northern Germany or the Alpine region instead of single grid points showed similar results.

## 4.3 SIMULATION RESULTS

First, we investigate the impact on the precipitation amount for different values of the correlation time scale $\tau$ and perturbation amplitude $\alpha_{ref}$. Corresponding domain

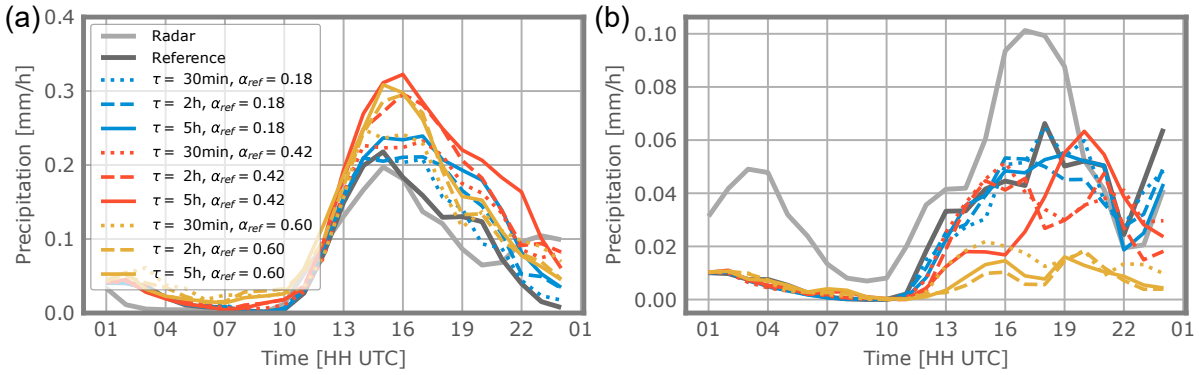


**Figure 23:** Domain averaged, hourly accumulated precipitation for a several different SSOSP settings. Left figure shows precipitation averaged over all grid points, that have a SSO-std$\geq 20\,m$, the right figure summarizes grid points with SSO-std$< 20\,m$.

averaged precipitation is displayed in Fig. 23. For orographic grid points increasing $\alpha_{ref}$ from 0.18 to 0.42 leads to more precipitation but, counter-intuitively, a further increase of $\alpha_{ref}$ causes a reduction in the precipitation amount. Interestingly, flat grid points only experience a reduction with increased $\alpha_{ref}$. One hypothesis for this decrease is that the perturbations enhance the mixing of boundary-layer and free tropospheric air, reducing instability, and leading to a reduction in precipitation. Monotonically increasing $\tau$ leads to more precipitation. More slowly evolving perturbations allow stronger updrafts to develop. These stronger updrafts cause an increased probability of convective initiation.

For maximum impact we chose $\alpha_{ref} = 0.42$ and $\tau = 5\,h$ for further ensemble experiments with 10 members. The domain averaged precipitation time series in Fig. 24 shows reasonable spread between ensemble members for both orographic and flat grid points. This spread is comparable to the spread generated by the original PSP scheme (see Kober and Craig (2016), Fig. 7d). While spread over orographic regions is desired, further insight is necessary to understand the behavior over flat grid points. Moreover, the behavior of individual ensemble members suggests a significant increase in precipitation over orographic grid points and a decrease over flat grid points. This is further illustrated in Fig. 24 where ensemble precipitation fields are displayed for two time steps: precipitation within the SSOSP scheme mainly occurs in the alpine region while most parts of Germany are precipitation free.

Overall, the SSOSP scheme shows a non-monotonic dependence on perturbation amplitude $\alpha_{ref}$, while increasing the correlation time scale $\tau$ appears to give increasing precipitation over orographic regions. For flat regions, however, the scheme causes an undesired reduction of precipitation.

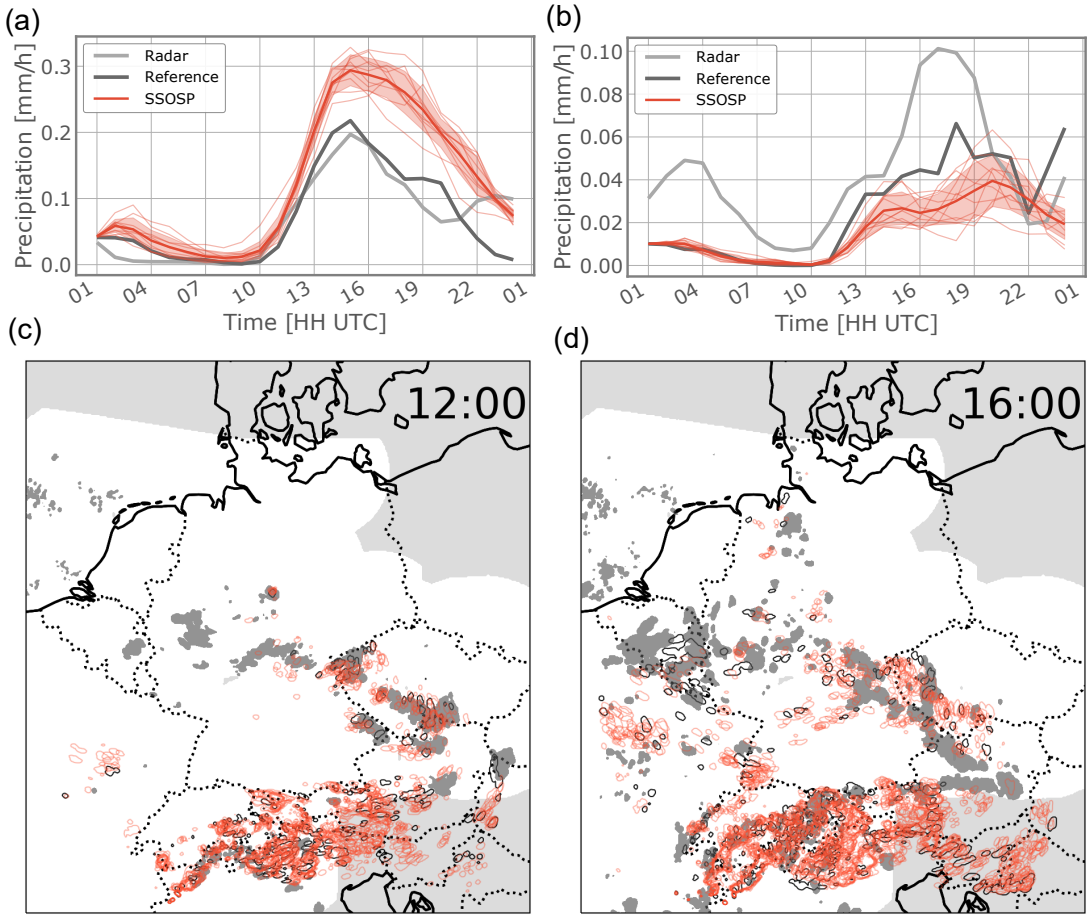


**Figure 24:** (a, b) Domain averaged, hourly accumulated precipitation for an ensemble simulation with $\tau = 5\,h$, $\alpha = 1.05$ ($\alpha_{ref} = 0.42$). The bold red line gives the ensemble mean, while red shading indicates the standard deviation. Single ensemble members are displayed in thin red lines. (a) shows precipitation averaged over all grid points, that have an SSO-std$\geq 20\,m$, (b) summarizes grid points with SSO-std$< 20\,m$. (c, d) Example precipitation contours at (c) 12 UTC and (d) 16 UTC for a 10 member ensemble. Shown are $0.1\,mm$ precipitation contours for radar in grey (filled), reference in black and SSOSP member in red.

## 4.4 SUMMARY AND DISCUSSION

We have developed SSOSP, a scheme to account for subgrid-scale orography and its impact on convective initiation. Doing so, we consider mechanical lifting by SSO taking into account the stratification of the atmosphere. Again, reasonable values of the tuning parameter $\alpha_{\text{tuning}}$ are of order one, supporting our physical reasoning.

The impact of the SSOSP on precipitation is not clear-cut. While we observe a precipitation increase over orographic regions, the scheme also leads to a reduction of precipitation over flat terrain. Furthermore, there is a non-monotonic relationship between the perturbation amplitude and the precipitation amount, which may be related to a combination of mechanical triggering and a mixing-induced reduction of CAPE. An ensemble simulation with *SSOSP* showed that spread over orographic terrain is generated but also over flat regions, which is conceptually not desired.

This behavior of the SSOSP makes it hard to employ. Furthermore, the impacts are not clearly complementary to those of PSP2: the precipitation impact of SSOSP is either too weak or coupled with an undesired impact on flat regions, and the generation of ensemble spread by SSOSP can also be achieved with PSP2. For these reasons, we refrained from combining the SSOSP perturbations and the PSP2 configuration. Conceptually, however, schemes for different subgrid-scale processes could easily be combined.

In the introduction, we referred to the open question of whether small-scale orography matters for convective initiation. Here we tested this by modeling subgrid-scale mechanical lifting on the smallest resolved scale. Our results suggest that the subgrid-scale orographic triggering effect most likely is not that relevant. On the one hand, to see the desired effect in mountainous regions, we had to increase the perturbation strength to a point where unwanted effects occurred over flat terrain. On the other hand, the perturbations introduced by our scheme are active on the smallest resolved scales of the km-scale model. Actual small-scale orography perturbations would be even smaller, and therefore likely even less influential. In general, our results confirm in a realistic setup, what has been observed in other studies (Tucker and Crook, 2005; Kirshbaum *et al.*, 2007b; Langhans *et al.*, 2013; Schneider *et al.*, 2018): conceptually, SSO may be relevant in regions of insignificant resolved orography; yet due to the fractal, scale-invariant structure of orography (Turcotte, 1987) regions of high subgrid-scale orographic variations are generally also regions of high grid-scale orography. This larger scale, resolved orography may dominate the triggering of convection so that SSO is virtually irrelevant for convective initiation.

# 5

# COLD POOL DRIVEN CONVECTIVE INITIATION (I): WHAT ARE CONVECTION-PERMITTING MODELS MISSING?

Now we focus on the role of cold pools in driving convective initiation and their representation in convection-permitting models. As introduced in Sec. 1.3.4, cold pools arise from evaporation in precipitating downdrafts and spread as density currents at the surface. Their gust fronts have the potential to trigger new convection mostly in the vicinity of already existing convection. Such secondary convective initiation can lead to the organization of convection, but also may be a pronounced trigger mechanism in the late afternoon and thereby contribute to the diurnal cycle of convection.

The difficulty of capturing the diurnal cycle of convection has been documented in a variety of km-scale models (Baldauf *et al.*, 2011; Hanley *et al.*, 2015; Clark *et al.*, 2016). In contrast to convection-parameterized models, where convection occurs too early during the day (Leutwyler *et al.*, 2017), convection in convection-permitting models starts too late during the day (Baldauf *et al.*, 2011; Clark *et al.*, 2016). This error can be improved by tuning the model, e.g., with the turbulent mixing length, but at the expense of other biases (Baldauf *et al.*, 2011; Schraff *et al.*, 2016; Leutwyler *et al.*, 2017). By using a more physically based approach, the PSP/PSP2 scheme (see Chapter 3), we were able to improve the onset of convection using stochastic perturbations to account for boundary-layer turbulence (see Chapter 3). However, a lack of precipitation in the late afternoon and evening remained. A similar behavior was documented by Rasp *et al.* (2018b) based on a 12-day summertime convective period over Germany. In addition, missing convective organization and structural deficits in cloud sizes were found (Rasp *et al.*, 2018b; Senf *et al.*, 2018; Panosetti *et al.*, 2019).

Given the deficits of convective forecasts from convection-permitting models, the question arises whether such models can simulate cold pool related triggering processes with sufficient accuracy. Grant and van den Heever (2016) have revealed a resolution dependence of cold pool gust fronts in idealized, two-dimensional LES simulations with grid sizes of 50-400m, namely weaker maximum vertical velocities in coarser models. This is consistent with studies showing a general sensitivity of vertical velocity to model resolution (e.g., Heinze *et al.*, 2017; Jeevanjee, 2017). Squitieri and Gallus (2020) recently identified a resolution sensitivity of cold pools in mesoscale convective systems with a higher frequency and larger areas in higher-resolution simulations. However, these studies have not investigated how well the triggering mechanisms are represented in convection-permitting models.

Considering the characteristics of cold pools, as described in the first paragraph (or Section 1.3.4), we expect cold pools to be relevant for the lack of late afternoon/evening precipitation and organization. A quantitative understanding of how cold pools and the associated circulations depend on the model resolution could potentially provide the basis for an improved treatment in km-scale models, similar to the improved repre-

sentation of marginally-resolved boundary-layer turbulence given by the PSP2 scheme from Chapter 3.

In this study, we address the following three research questions:

1. Are cold pools a significant factor for the diurnal cycle and the organization of convection for mid-latitude, continental weather?

2. Can convection-permitting models adequately represent cold pool driven convective initiation processes?

3. If not, which misrepresented processes are responsible for the deficiencies in cold pool driven convective initiation?

Since observations of cold pools are sparse and often incomplete, we base our work on high-resolution simulations of realistic weather situations over Germany. Simulations with 156 m grid sizes are available from the HD(CP)$^2$ project (Heinze *et al.*, 2017). The simulations have been thoroughly compared to observations, confirming that their turbulence profiles or cloud distributions are sufficiently comparable to observations (Heinze *et al.*, 2017). We assume that cold pool dynamics will be sufficiently resolved as well. The 156 m simulation can be compared with additional simulations at 312 m and 625 m grid spacing. By comparing cold pool and convective initiation characteristics between these resolutions, we draw conclusions regarding the ability of convection-permitting models to represent cold pool driven convective initiation. In doing this, we assume that higher resolution produces more realistic cold pools and cold pool driven convective initiation and that the simulations with 625 m grid spacing are sufficiently representative of convection-permitting simulations - or at least its higher resolution end of the spectrum. Due to the immense computational cost of the high-resolution simulations, we are only able to analyze four summertime convective days. These days have been selected with a range of large-scale meteorological conditions that we believe are representative of the diversity typically found in nature. To address the question of which processes are responsible for the differences between the simulations at different resolutions, we apply linear causal effect estimation based on Pearl's causality (Pearl, 2009) and the graphical representation of causal structures (Wright, 1921; Pearl, 2013; Chen and Pearl, 2015). This enables a formalized estimation of different direct or indirect causal effects in the model data. While this framework is not yet commonly applied in atmospheric sciences, some climate studies have applied and developed related causality concepts (e.g., Ebert-Uphoff and Deng, 2012; Runge *et al.*, 2015; Kretschmer *et al.*, 2016; Samarasinghe *et al.*, 2019). A recent review of causality inference in Earth system science can be found in Runge *et al.* (2019).

This chapter is structured as follows. In Section 5.1 we lay out relevant details of the numerical simulations and the days to be investigated, how we identify cold pools, cold pool boundaries and convective initiation, and the applied diagnostic and causal methods. Section 5.2 presents the results of this study organized around the three research questions. In Section 5.3, we provide a summary of our results and discuss their underlying assumptions and implications.

## 5.1 DATA AND METHODS

### 5.1.1 *ICON-LEM simulations*

The basis of this study is a set of high-resolution summertime simulations over Germany performed with the large-eddy version of the ICOsahedral Non-hydrostatic atmosphere model (Zängl *et al.*, 2015; Dipankar *et al.*, 2015) in limited area model configuration (Heinze *et al.*, 2017) as part of the HD(CP)$^2$ project. Simulations have been performed at horizontal resolutions of 156 *m*, 312 *m* and 625 *m* on a number of days with a range of meteorological conditions. Details of the model setup and a comprehensive evaluation of the simulations can be found in Heinze *et al.* (2017). The meteorological situations, thermodynamic conditions and cloud and precipitation properties seen in observations are generally well captured in the ICON-LEM simulations and show improvements when compared to km-scale simulations (Heinze *et al.*, 2017). For comparability, the model output is interpolated onto a common 1.2 km latitude-longitude grid for all three resolutions. We select a region of 5.503-14.496° E and 47.599-54.496° N for our analysis, which is available for all dates and resolutions. The simulation output is available from 6:00 UTC to 24:00 UTC with 5-minute output frequency (simulations start at 00:00 UTC).

### 5.1.2 *Selected days and their synoptic situations*

We have selected four days that represent diverse large scale conditions and exhibit summertime convection and cold pools, namely 5 July 2015, 29 May, 6 June and 1 August 2016. Snapshots of buoyancy intensity (as defined in Sec. 5.1.3.1), precipitation and cold pool boundaries are shown for each day in Fig. 25. Time series of selected domain aggregated variables that capture the large scale convective situation are displayed in Fig 26.

On 5 July 2015, Germany lay in the warm sector of a frontal system that had approached from France. The approaching front was led by a convergence line, and relatively strong south-westerly winds were present. High CIN, CAPE, bulk Richardson number (BRN) [8] and significant wind shear were present over the simulated domain (see Fig. 26a, b, e, f). Squall line-like structures developed in the simulations, traveling over Germany from West to East, accompanied by two enormous cold pools (see Fig. 25a, the second enormous cold pool is entering the domain from the West at this time). On 29 May 2016, a low-pressure system connected to an omega block was centered over Germany. Cyclonic rotation, large scale lifting (Fig. 26d) and intense and damaging flooding, especially in Braunsbach, Baden-Württemberg, occurred. Several cold pools are found in the simulations. The high-pressure system of the omega block slowly shifted subsequently to fully enclose Germany on 6 June 2016, suppressing large scale lifting (Fig. 26d) and wind shear (Fig. 26f), but leaving high values of CAPE and BRN (Fig. 26b, e). During the day, widespread scattered convection is simulated, and

8 The bulk Richardson number is computed as specified by Markowski and Richardson (2011). For the 0-6 km mean wind, the model levels 76 to 150 are used and for the 0-500 m mean wind model levels 139 to 150 are used.

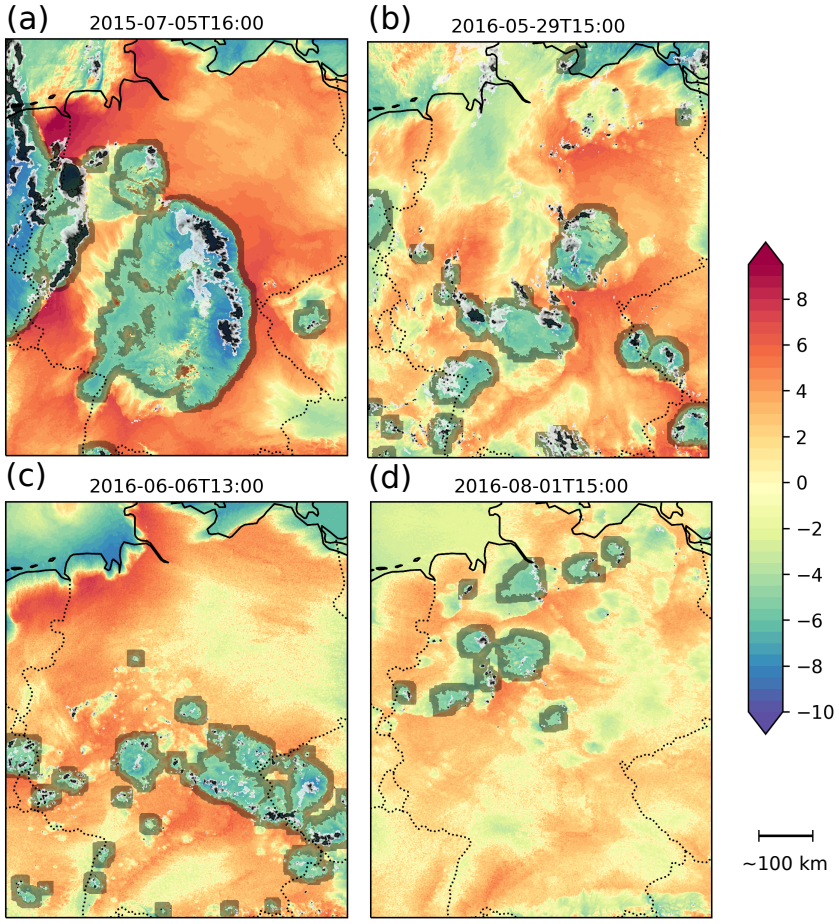


**Figure 25:** Example snapshots of buoyancy intensity (see Sec. 5.1.3.1) [$m\,s^{-1}$] in color, cold pool boundary regions in grey shading and precipitation intensity in white ($1\,mm\,h^{-1}$) to black ($20\,mm\,h^{-1}$) are displayed for each day from the 156 m grid size simulations. Dotted and solid black lines show country borders and coastlines respectively.

an organized band forms from merged cold pools and spreads over South-Eastern Germany during the afternoon. Due to the intense convective activity and high variability in underlying forcing, the period spanning 29 May and 6 June 2016 has been considered in a number of studies on convection (Rasp *et al.*, 2018b; Baur *et al.*, 2018; Bachmann *et al.*, 2019; Keil *et al.*, 2019) and analyzed in detail by Piper *et al.* (2016). As explained in Chapter 2, these days are also considered in chapters 3, 4 and 6. On 1 August 2016, a low-pressure system over Scandinavia and a small high-pressure system over Switzerland dominated the synoptic situation over Germany. Relatively small CAPE values (Fig. 26b), a deep boundary layer, some large scale subsidence, and significant shear are simulated (Fig. 26c, d, f). The simulations produce many small cold pools over northern Germany (see Fig 25d).

### 5.1.3 *Detection of cold pools and their edges*

#### 5.1.3.1 *Cold pool detection*

We identify cold pools using the density potential temperature perturbation and precipitation on the lowest layer above the surface. Density potential temperature is defined as $\theta_\rho = \theta(1 + 0.608 r_v - r_w - r_i - r_r - r_s - r_g - r_h)$ (Emanuel, 1994, eq. 4.3.2 and 4.3.6, using an approximation to first order in $r$ analogous to eq. 4.3.1). Here $\theta$ denotes the

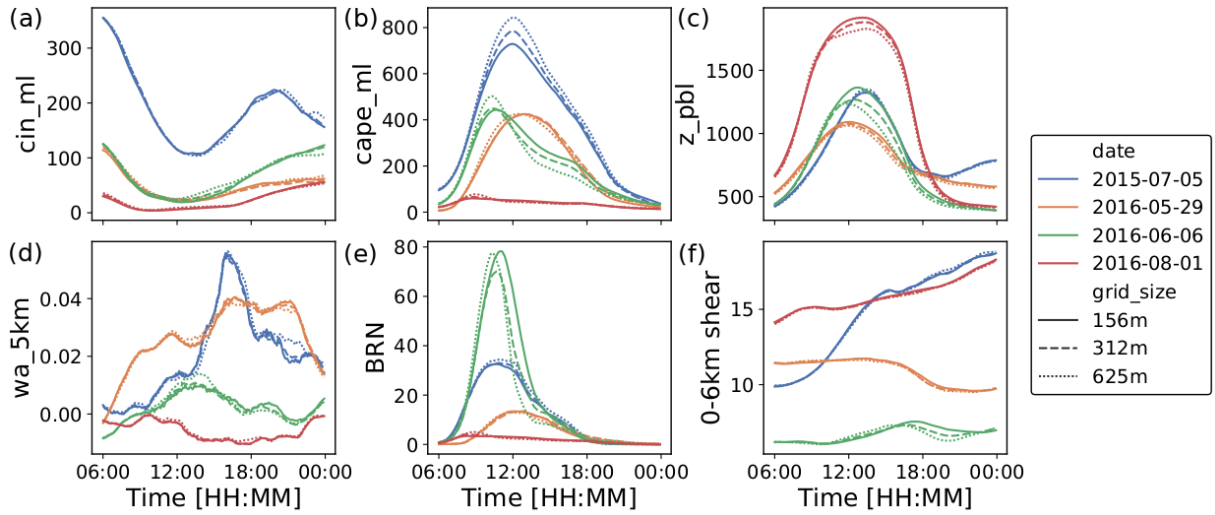


**Figure 26:** Time series of domain aggregated variables describing the synoptic situation: a) mean layer convective inhibition [J $kg^{-1}$] (direct model output) ; b) mean layer convective available potential energy [J $kg^{-1}$] (direct model output); c) height of the planetary boundary layer [m ASL] (direct model output); d) vertical velocity at approximately 5 km [m $s^{-1}$]; e) Bulk Richardson number (BRN) (computed according to Markowski and Richardson, 2011); f) 0-6km wind shear [m $s^{-1}$] (model levels 150 and 76 are used). For BRN, the domain median has been computed due to the occurrence of high outlier values. Otherwise, domain averages are displayed.

potential temperature, $r_v$ the water vapour mixing ratio and $r_w$, $r_i$, $r_r$, $r_s$, $r_g$ and $r_h$ the mixing ratios of liquid cloud water, cloud ice, rain, snow, graupel and hail, respectively.

We calculate the local perturbation in density potential temperature, $\theta'_{\rho,0}$, relative to a moving average, $\overline{\theta_\rho}^m$, of filter size 166 pixels ($\approx 200\,km$) horizontally and $8\,h$ in time. The moving average is computed with reflection at the boundaries. This combination of spatial and temporal filtering enables the identification of large cold pools (especially on 5 July 2015) while background $\theta_\rho$ gradients at the coast are still sufficiently resolved to detect cold pools there. A small sensitivity analysis with different spatial filtering scales and with/without temporal filtering showed no strong influence on the qualitative behavior of our results. We further calibrate the perturbations by subtracting the domain-mean density potential temperature perturbation: $\theta'_\rho = \theta'_{\rho,0} - \overline{\theta'_{\rho,0}}$. We then identify regions where the density potential temperature perturbation $\theta'_\rho$ is below $-2\,K$ and apply the watershed-merge segmentation method of Senf *et al.* (2018) to identify cold pools. The segmentation algorithm partitions these cold areas into cold pool objects by first shrinking the objects to identify object cores and assigning each edge pixel to the nearest core. If part of the interface between two objects thus found is farther from the environment than half of the maximum distance to the environment in one of the objects (i.e., the interface passes through an inner part of the object), they are merged again, otherwise (if the interface lies close to the outer edge), they are kept separate (Appendix A in Senf *et al.*, 2018). Cold pool objects smaller than 20 pixels, i.e., with an equivalent diameter of $\approx 6\,km$, are excluded. Following Senf *et al.* (2018), this criterion has been chosen to only include cold pools with an equivalent diameter larger than the effective model resolution of the coarsest model ($\approx 5\,km$; see Heinze *et al.*, 2017). In

order to exclude cold areas like sea breezes or valley inversions that are not connected to convection, we stipulate that precipitation in at least one pixel of the cold pool object must exceed $10\,mm/h$, and discard objects otherwise.

This method successfully identifies roughly convex cold pool areas connected to convective precipitation. It is not very sensitive to the choice of thresholds, and is more robust than more complex derived cold pool identifiers suggested by Drager and van den Heever (2017). Given our selected criteria, however, we will not be able to identify very weak cold pools (with regard to buoyancy anomaly), very small cold pools or non-/weakly-precipitating cold pools, which may especially exclude very young and old cold pools. This may have some implications for our results, but due to the small size and the weak strength of young and old cold pools, we expect to have captured most cold pools that are relevant for convective initiation. The expectation that dissipating cold pools trigger fewer new convective cells is also supported by results from Fournier and Haerter (2019).

For each air column, we define the cold pool *intensity* $I$ using the integrated buoyancy anomaly $b_{int}$ over the lowest five layers:

$$I = sign(b_{int})\sqrt{|2b_{int}|},\ \text{with}\ b_{int} = \int_{0\,\mathrm{m}}^{\approx 150\,\mathrm{m}} b\,dz\ \text{and buoyancy}\ b = g\frac{\theta'_\rho}{\overline{\theta_\rho}^m}.$$

This definition is chosen to agree with the cold pool intensity used by Feng *et al.* (2015), except for the sign which we choose to be positive for positively buoyant air (and therefore negative within cold pools). We choose fixed model layers as cold pool depth for simplicity. The lowest 5 model layers (ca. 150m) were selected as they include the buoyancy perturbation of most cold pools - integrating slightly higher than the cold pool top does not affect the result, since the buoyancy anomaly is small at these levels. Fig. 25 shows snapshots of intensity for a selected time on each of the four days.

The algorithm for this cold pool detection was provided by R. Heinze and modified in collaboration with S. Schäfer.

#### 5.1.3.2 *Cold pool boundary regions*

It is not the cold pool itself but the leading gust front with thermodynamic or mechanical lifting that drives the initiation of new convection. Detecting the gust fronts is however not trivial in an automated framework and associating it to a specific cold pool is also ambiguous. For instance, using horizontal wind speed or vertical velocity above a threshold to identify cold pool gust fronts will also capture other phenomena like land-sea breezes, orography or convergence lines. Recently, Fournier and Haerter (2019); Henneberg *et al.* (2020) developed two approaches for detecting cold pool gust fronts based on horizontal wind speeds, which may provide better alternatives in the future. By using the above cold pool definition, the gust fronts tend to lie just outside of the identified cold pools (see Fig. 27). To avoid the difficulties of objectively identifying gust fronts, we define a cold pool boundary region as a zone around each cold pool with $\approx 25\,km$ width, which is the region most likely to contain the gust front. With this definition, large buoyancy gradients and forced ascent will occupy only a fraction of this region. To characterize the gust front itself, we consider 95th percentile values,

rather than averages over the entire gust front region. An advantage of considering this cold pool boundary region is that the thermodynamic lifting will be included, which may not necessarily be located directly at the gust front. Grid points belonging to the original cold pool or other cold pools are excluded from the boundary region. Overlap with other cold pool boundary regions is, however, possible. We will consider overlapping boundary regions as a proxy for cold pool collisions.

Comparing cold pool boundaries with vertical velocity in Fig. 27, many line structures of strong vertical velocity can be identified as cold pool gust fronts (e.g., A and B in Fig. 27b) and generally they are positioned somewhere within the defined cold pool boundary region. This confirms that such cold pool boundaries are a useful proxy for their gust fronts.

### 5.1.4 *Defining convective initiation*

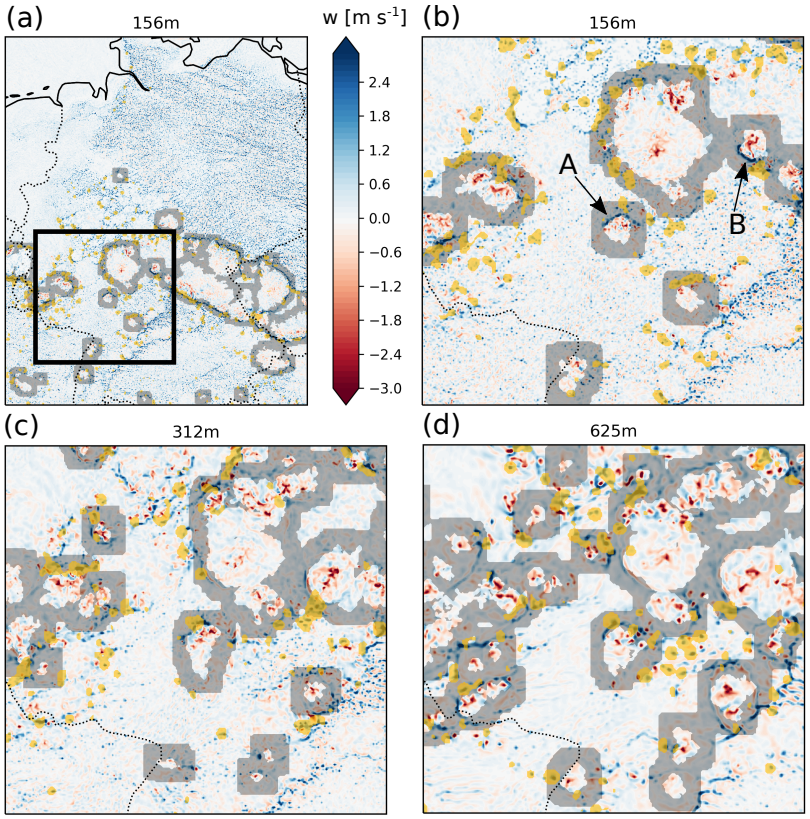


**Figure 27:** Example snapshots of vertical velocity at $\approx 1\,km$ in color $[m\,s^{-1}]$, cold pool boundaries in grey shading and locations of convective initiation in yellow for each model resolution on 6th June 2016 13:00 UTC. The black box in (a) marks the displayed region for (b), (c) and (d).

There is no standard definition for identifying convective initiation. In this study, we aim to relate the initiation of a convective cloud to properties in the boundary layer, such as the gust front vertical velocity. Hence, to obtain an independent identification of vertical velocity at the gust front and convective initiation, we refrain from using vertical velocity as an indicator for convective initiation. We also avoid using quantities such as precipitation particles, which occur later in the convective cloud life cycle and may be difficult to associate with boundary-layer features due to advection of the cloud to neighboring grid points. With these conditions in mind, we choose the decisive variable

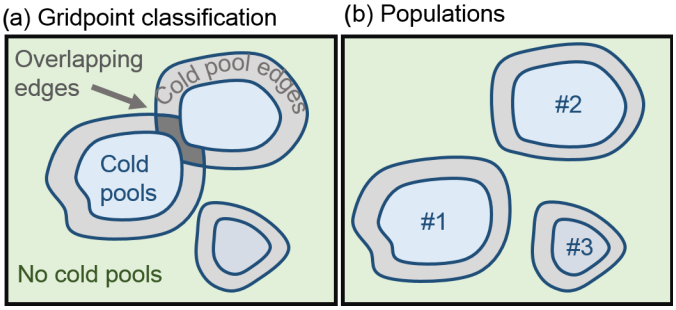


**Figure 28:** Schematic illustration of (a) classification of gridpoints into cold pools, cold pool boundaries and no cold pools; and (b) cold pool populations.

for convective initiation to be the temporal change of cloud water content in the mid-troposphere. The onset of deep convection is associated with a rapid generation of cloud droplets extending through a large part of the tropospheric depth; hence large values of the temporal change of cloud water content $\frac{\partial r_w}{\partial t} = \partial_t r_w$ are associated with convective initiation. The following definition for convective initiation is used:

$$r_w^* = G\left[\sum_{z=2.5\,km}^{8\,km} \partial_t r_{w_z}\right] > b_{r_w^*}.$$

The temporal change is computed offline with $5\,min$ time intervals. To eliminate the effect of high cirrus clouds and low stratocumulus, we consider only mid-tropospheric levels and compute the sum of $\partial_t r_w$ over the model levels 60-110, corresponding approximately to 2.5-8 km. To reduce noise and to allow for some spatial displacement, we apply a Gaussian filter $G$ using two grid boxes as the standard deviation for the Gaussian kernel. Convective initiation is determined to have occurred at grid points where $r_w^*$ exceeds a threshold $b_{r_w^*}$ that depends on day and resolution. The threshold value is defined as the $\alpha$-percentile of the $r_w^*$ field, where $\alpha = 1 - f_{prec}$ with $f_{prec}$ being the relative frequency of precipitating grid points for that day and model resolution. A grid point is identified as precipitating if the rain rate exceeds $1\,mm\,h^{-1}$. In this way, we identify as many grid points as having convective initiation as we have precipitating gridpoints for that day and model resolution. We note that single grid points can be identified as convective initiation for subsequent time steps as long as the criterion holds.

Example fields of convective initiation are displayed in Fig. 27. Further visual comparisons with precipitation fields (not shown) reveal that most identified convective initiation is indeed related with an onset of precipitation at later times and that advection of deep convection is only rarely identified as convective initiation.

### 5.1.5 *Diagnostic approaches*

GRID POINT CLASSIFICATION With the criteria defined above, we can partition the horizontal domain into four categories, namely 1) *no cold pools*, 2) *cold pools*, 3) *cold pool edges*, and 4) *overlapping cold pool edges*, as schematically illustrated in Fig. 28a. Overlapping cold pool edge regions are also included in the cold pool edge category, as discussed below.

**Table 2:** Quantities that represent essential characteristics of cold pools and their gust fronts.

| Variables | region | description |
|---|---|---|
| Intensity | cold pool | Buoyancy anomaly $\theta'_\rho$ integrated vertically over the 5 lowest model levels ($\approx 150\,m$). Area averages (mean) or integrals (sum) are considered. |
| D-mass flux | cold pool | Downward mass flux, i.e., negative values of $\rho w$ at ca. 1km, integrated over each cold pool area (sum). |
| Radius | cold pool | Equivalent radius, estimated as the radius of a circle with the same area as the cold pool. |
| Precipitation | cold pool | Precipitation inside cold pools, integrated over the cold pool area. |
| Gust front vertical velocity | edges | 95th percentile of vertical velocity in the cold pool boundary region at ca. 1km height. This height was visually identified as the height of most pronounced gust fronts. |
| Gust front mass flux | edges | Upward mass flux inside the cold pool edge regions at 1km height. Both the area integrated (sum) and area average (mean) are used. |
| Buoyancy | edges | 95th percentile of buoyancy at ca. 150m height. Potential temperature, moisture or surface fluxes are not considered, as buoyancy is expected to hold most of their information relevant for convective initiation. |
| Triggering probability, $P[CI]$ | edges | Number of gridpoints with convective initiation divided by the number of grid points in the cold pool edge region. |

COLD POOL OBJECTS We can further identify every single cold pool as an object and compute cold pool characteristics averaged or integrated over its cold pool region (see Fig. 28b). To each cold pool, we associate a cold pool edge, which represents the $25\,km$ wide boundary region defined above. Recall that grid points belonging to a cold pool are excluded from the cold pool edge. Regions that overlap with another cold pool's edge region are included. This means that some grid points may be counted as part of the cold pool edge region for more than one cold pool. Again, for each cold pool edge, area-averaged or -integrated quantities can be computed, but also 95th percentiles to represent the more extreme behavior of the actual gust fronts. As defined in Table 2, eight quantities are used to describe the essential characteristics of cold pools and their boundaries. For the cold pool, these are intensity, downdraft mass flux, radius and precipitation, while for the cold pool edge region, we use vertical velocity, mass flux, and buoyancy. The relative frequency of grid points where convective initiation occurs defines the probability of convective initiation, $P[CI]$, or *triggering probability*.

COLD POOL POPULATIONS Given these characteristics, we can compare the distributions of cold pool properties (populations) between model resolutions. For this purpose, we will generally show the median of the cold pool distributions and vertical bars for the 95% confidence interval, which is computed via bootstrapping with 1000 draws. We

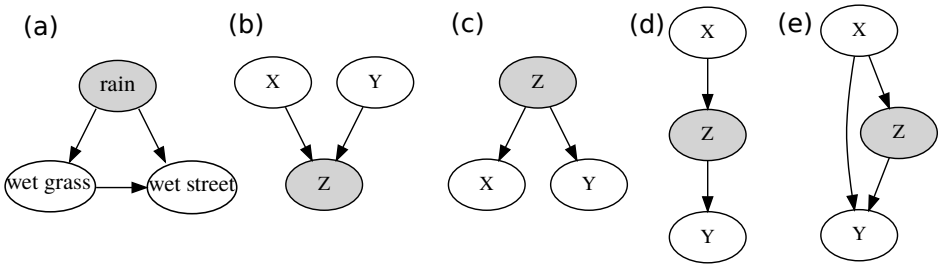


**Figure 29:** Example directed, acyclic graphs illustrating the concepts of a collider, d-separation or the single door criterion. See text for details.

note that cold pools from 5 min output are used and that the same, temporally correlated cold pools appear more than once in the cold pool populations. Hence the data is not independent, and the significance of our results will likely be overestimated. This possibility has been briefly tested and confirmed by sub-sampling the data every full hour.

### 5.1.6 *Estimating linear causal effects*

To determine the processes through which model resolution affects cold pool driven convective initiation, we apply linear causal effect estimation. In general, associational expressions like regression estimates do not prove causal relationships – correlation is not causation[9] – because common drivers may produce associations between two variables without one causing the other. Furthermore, in complex systems, there may be a chain of causality through different variables, and it is of interest to quantify and compare different pathways from the original cause to the effect, since they may represent different physical mechanisms. To do this, we follow the framework reviewed by Chen and Pearl (2015) and Pearl (2013, 2009) based on Pearl's causality (Pearl, 2009) and the graphical representation of causal structures, first proposed by Wright (1921). This framework is based on the causal graph of the underlying phenomenon, which has to be determined using a priori process understanding. We will assume that all variables are linearly related, so that multiple linear regressions can be used to determine single causal effects. The graphical analysis can also be applied for nonlinear relationships, but for this work, we will avoid the additional complexity. Importantly, the required predictors for the multiple linear regressions are selected based on the structure of the causal graph by the so-called *single door criterion*. The remainder of this section will more carefully define a causal effect, and illustrate the analysis method with some simple examples of causal graphs. To define the single door criterion, we will need to introduce additional concepts, including directed acyclic graphs, collider and d-separation. Note that these latter concepts are not needed to understand the application of causal analysis to our cold pool experiments in Section 5.2.3, provided that the reader is willing to accept the structure of the multiple linear regressions that results from the analysis.

9 But they are correlated.

For a more thorough introduction to linear causal analysis, we refer the reader to Chen and Pearl (2015).

A **causal effect** of $X$ on $Y$ can be defined as the change of the expected value of $Y$ when $X$ is changed actively (by intervention) from $X = x$ to $X = x + \Delta x$. The active intervention is crucial because it eliminates any factors that otherwise impact $X$. Consequently, randomized experiments or model sensitivity studies are commonly performed to quantify causal effects. For example, $X$ can be the wetness of grass and $Y$ the wetness of the adjacent street. If one wants to test the hypothesis that the wet grass causes the wet street (see Fig. 29a), simply computing the correlation of the wetness of the street with the wetness of the grass will give a biased estimate, because the common cause of rain has not been included. One way to avoid such a spurious correlation is by active intervention. If the grass is actively moistened, e.g., by using a sprinkler, the correlation with a wet street will be reduced.

Fortunately, under certain circumstances, it is possible to determine causal effects solely from associations, without doing any intervention. In the simple example of wet grass and street, the solution would be to condition on the presence of rain, e.g., by including rain as an additional predictor in a linear regression. On rainy days, everything will be wet, but on dry days, the wetness of the grass and street is determined by sprinklers and street cleaners and will not be correlated, thus reducing the overall spurious correlation. For more complex systems, it can be difficult to determine a set of parameters that can be conditioned on to allow a causal connection to be identified, or even if such a set of parameters exists. Here we identify these circumstances with the so-called *single door criterion*, which makes use of the following concepts.

**Directed acyclic graphs** are graphs with directed edges and no cyclic structures. They are powerful tools to visualize causal structures and enable graphical criteria to estimate causal effects. Example graphs are shown in Fig. 29. In this application, the nodes of a directed acyclic graph represent variables of a causal model, which are causally influenced by their parent nodes and themselves cause changes in child nodes. Parents and children of a node are determined by the direction of the arrows connecting them: arrows pointing towards a node emanate from its parent nodes and arrows emanating from a node point to its child nodes. In the example of Fig. 29b, Z is a child of both X and Y, whereas in Fig. 29c, Z is a parent of X and Y. The concepts of ancestors and descendants of a node follow naturally. By assuming linearity, a directed edge connecting a parent and a child node can be associated with single *path coefficient* representing the direct causal effect, i.e., the rate of change of the child variable, when the parent is actively perturbed. It is these path coefficients that we wish to estimate by appropriate multiple linear regressions.

A **collider** is a special element on a path in a causal graph, where two arrowheads collide at one node. If a path between X and Y passes through a collider, that path is said to be *blocked*: despite its connection by the colliding edges, no information will pass through this pathway. In Fig. 29b, for instance, Z is a collider for the path connecting X and Y through Z. Collider nodes will have important consequences, as described in the following paragraph.

Directed separation, or **d-separation**, encapsulates the graphical conditions that correspond to a missing causal relationship. There are three rules for two variables to be d-separated (Chen and Pearl, 2015):

1. If there is no *active* path between X and Y, the variables X and Y are d-separated. An active path refers to any path that is not blocked by a collider. In Fig. 29b X and Y are d-separated as the only path is blocked by the collider Z. In Fig. 29c, d and e, X and Y are not d-separated as active paths exist.

2. X and Y can also be d-separated *conditioned on a set of nodes Z*, if there is no active path between X and Y without traversing Z. Meaning, if Z is included in all active paths from X to Y, X and Y are d-separated conditioned on Z. In Fig. 29c and d, X and Y are only d-separated if conditioned on Z, as the only active path between X and Y is via Z.

3. If a conditional set Z contains a collider or a child of a collider, X and Y cannot be d-separated conditioned on Z. Hence, X and Y in Fig. 29b are not d-separated conditioned on Z.

If two variables X and Y are d-separated conditioned on a set of nodes Z, the partial correlation coefficients between X and Y, where Z is accounted for, vanish (see, e.g., Chen and Pearl, 2015). By applying this concept, causal graphs can be linked to statistical (conditional) independence.

With these concepts, we can now state the conditions under which we can determine causal effects from non-interventional data. In a linear framework, this is formally given by the **single door criterion**. It states that the direct causal effect of X on Y within a causal graph G is *identifiable* if a set of variables Z exists with the following criteria: (i) Z contains no descendent of Y, and (ii) X and Y are d-separated conditioned on Z in the hypothetical graph G' where the direct path from X to Y is removed (this implies that Z does not contain a collider or a child of a collider). Most importantly, such identifiable path coefficients are equal to the regression coefficient that is obtained from multiple linear regression predicting Y using not just X but also Z as predictors. We will refer to Z as the *adjustment set*. The aggregated causal effects of parallel or sequential paths can then be derived from single path estimates by addition or multiplication, respectively.

Considering the example graph in Fig. 29e, Z d-separates X from Y in the graph where the direct X-Y path is removed (here corresponding to Fig. 29d). Hence the direct effect of X on Y is identifiable and can be estimated via the regression estimate of X from a linear regression where both X and Z are used as predictors, and Y is the predictand. To estimate the direct causal effects between X and Y in Fig. 29b, c and d (which are all zero due to missing, direct links) Z has to be included in Fig. 29c and d, whereas it must not be included in Fig. 29b. Returning to the example in Fig. 29a, the single door criterion shows that, to estimate the causal effect of the wetness of grass on the wetness of the street from observational data (no intervention), the rain amount has to be included in the linear regression. [10]

10 The single door criterion is sufficient, but not a necessary criterion to identify causal effects. Further criteria, the backdoor criteria and instrumental variables, enable additional identification (Chen and Pearl, 2015). In our case, however, the single-door criterion is adequate. It is also possible to identify different

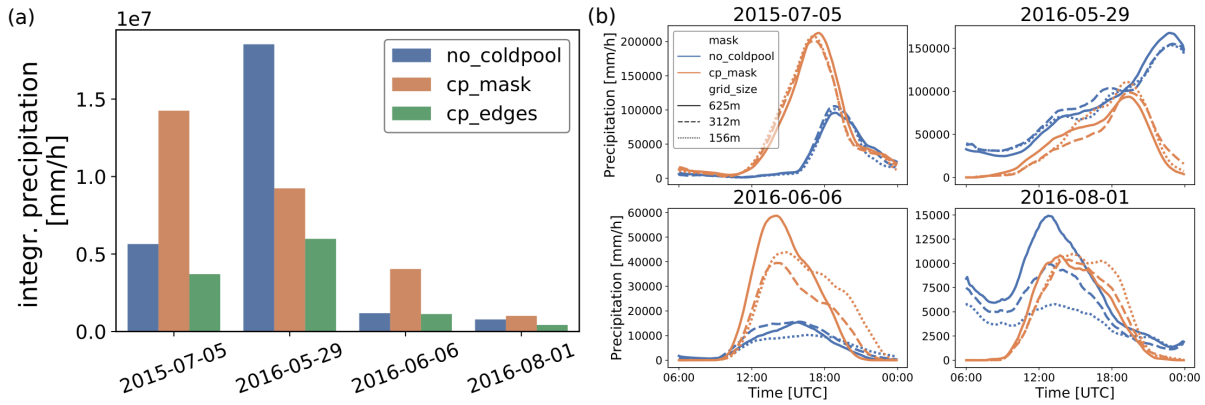


**Figure 30:** a) Domain and time integrated precipitation for the 156 m resolution simulation outside cold pools, within cold pools and within cold pool boundaries. b) Time series of domain integrated precipitation within cold pools and outside cold pools. A 3 h running mean has been applied to smooth the data.

In summary, this framework enables a graphical criterion, namely the single-door criterion, which specifies if and how causal conclusions can be drawn from purely statistical associations. The causal, graphical structure, however, must be provided using a priori understanding of the governing processes. We apply this framework to identify the impact of model resolution on cold pools and convective initiation in Section 5.2.3.

## 5.2 RESULTS

### 5.2.1 *Linking precipitation and convective initiation with cold pools*

The first question we address is the relevance of cold pools for precipitation and especially for its diurnal cycle and organization. We focus on the 156 m grid size simulations as the qualitative results are mostly similar for the other two resolutions. We compare temporally and spatially integrated precipitation amounts inside cold pools (*cp_mask*) and outside cold pools (*no coldpool*) in Fig. 30a for each day using the gridpoint classification described previously. First, we consider some general behavior of precipitation inside and outside of cold pools on these four days. The magnitude of total precipitation clearly depends on the day, with 29 May 2016 showing the highest total amplitude and 1 August 2016 the lowest. But the relative importance of precipitation inside cold pools also varies: on 6 June 2016 precipitation inside cold pools (*cp_mask*) was approximately three times greater than outside cold pools whereas on 29 May 2016 precipitation outside cold pools dominates. Nonetheless, precipitation inside cold pools contributes substantially for all days, ranging from approximately 50 to 300 percent of precipitation outside cold pools. This corroborates that precipitation predominantly produces cold pools and indicates the potential importance of cold pools. Fig. 30a also shows that significant amounts of precipitation occur in cold pool edge regions, which provides a first indication of the role of cold pools in initiating convection.

indirect causal effects for nonlinear or non-parametric models (Pearl, 2014). The necessary conditions for identification are however more restrictive and the required estimation methods more complex.

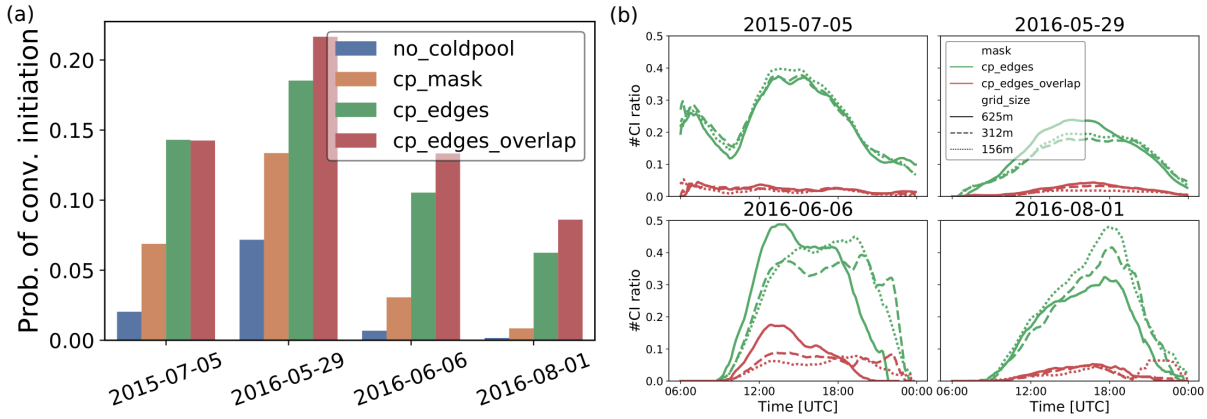


**Figure 31:** a) Triggering probability in the four grid point categories. b) Ratio of convective initiation inside cold pool boundaries or overlapping cold pool boundaries to total convective initiation for each day and model resolution. A 3 h running mean has been applied to reduce noise.

Time series of domain integrated precipitation for the different model resolutions are shown in Fig. 30b. Again, we find a strong dependence of the diurnal cycle on the synoptic conditions. In general, cold pool related precipitation, but also *no coldpool* precipitation, tends to be more dominant during the day and decreases in the evening. Also, the phase and amplitudes of the diurnal cycles do not completely coincide with the diurnal cycle of precipitation outside cold pools. On 5 July 2015, when two enormous cold pools develop, cold pool precipitation increases until the early afternoon without noteworthy precipitation outside cold pools. On 29 May 2016, the increase in precipitation during the day occurs both inside and outside cold pools. However, precipitation outside cold pools continues to increase throughout the night, likely related to synoptically driven ascent. Interestingly, on the other two days, the rapid evening decline of cold pool related precipitation tends to occur later than for precipitation outside of cold pools, again suggesting a role for cold pools in maintaining convection.

We now focus on the connection between cold pool gust fronts and convective initiation, which indicates the influence of cold pools on precipitation. Fig. 31a shows the triggering probability as function of gridpoint classification and day. In *no_coldpool* regions, we find the lowest probability of convective initiation for all days. Triggering is substantially enhanced inside cold pools. Note that this behavior is based on our definition of convective initiation as a strong increase in cloud water content and several possible explanations for convective initiation inside cold pools exist.[11] As expected, the cold pool edge regions have even higher triggering probability. Depending on the day, from 7% to almost 20% of gridpoints inside cold pool boundaries initiate convection. The highest triggering probability on most days is found for overlapping edges, with an increase in the probability of CI of approximately 3% in comparison to all cold pool

11 These explanations include 1) merged cold pools which may still have active gust fronts within the identified cold pool regions; 2) gust fronts at the edge, but still inside identified cold pool region; 3) rapid growth or propagation (not advection) of existing cells within the cold pool. A further reason could be due to the wrong identification of advected cells.

edges. These findings are consistent with previous studies showing that cold pools are very efficient at initiating new convection nearby.

To estimate the absolute impact of cold pools on total convective initiation, Fig. 31b displays the diurnal cycle of convective initiation within cold pool edges and overlapping edges relative to total convective initiation. We find that triggering by cold pool edges accounts for up to 50% of total triggering. For the last two days, this is especially strong in the afternoon and evening. Although we have seen a high efficiency in triggering by overlapping edges, their contribution to the total convective initiation is small (Fig. 31b), being usually well under 10%.

The results address the first research question, showing that cold pool driven convective initiation plays a substantial role in the diurnal cycle and the organization of convection. Strong links between precipitation and cold pools are observed, with cold pool triggering contributing up to 50% of total convective initiation with particular relevance in the evening in some synoptic situations. Cold pool boundaries are found to be particularly important for triggering of convection. In two cases, this impact is felt strongly during the evening, when non-cold pool triggering has decreased in strength.

### 5.2.2 *Sensitivity of cold pool properties to resolution*

Having established strong indications for the influence of cold pools on the initiation and diurnal cycle of convection, it follows that the appropriate representation of cold pool dynamics in convection-permitting models is vital. We will evaluate whether cold pools and the related convective initiation are indeed appropriately represented at all resolutions and hence insensitive to model resolution.

For the first two days in Fig. 30b no systematic differences in the diurnal cycle of precipitation between model resolutions are visible. For the third case, 6 June 2016, the coarsest model shows a higher peak in the early afternoon, and increased model resolution results in a shift of cold pool related precipitation towards the evening. A similar, but less pronounced behavior is also visible for the last day, 1st August 2016. Surprisingly, on this day, precipitation outside cold pools shows a relatively strong sensitivity to model resolution, related to many small, scattered precipitation cells over Northern Germany in the coarsest model, which become less frequent at higher resolutions.

To investigate the resolution dependence of cold pools more systematically, we now examine distributions of cold pool properties. Fig. 32 summarizes the most evident differences between resolutions. For all four days, the number of cold pools is reduced at higher resolutions. The size of cold pools (i.e., the equivalent radius), on the other hand, increases with resolution. These observations are also in accordance with the visual impression given in Fig. 27. Cold pools are generally more intense (integrated or averaged over their areas), and have stronger precipitation (mainly on first and last day) and integrated downward mass flux at higher resolutions. Some of these differences between model resolutions are relatively small compared to the day-to-day variability. Especially the first day (5 July 2017) is different since it has only a few, enormous, cold pools (see Fig. 25). Overall, the dependencies of cold pool properties are quite systematic, despite the large differences in their absolute values from day to day.

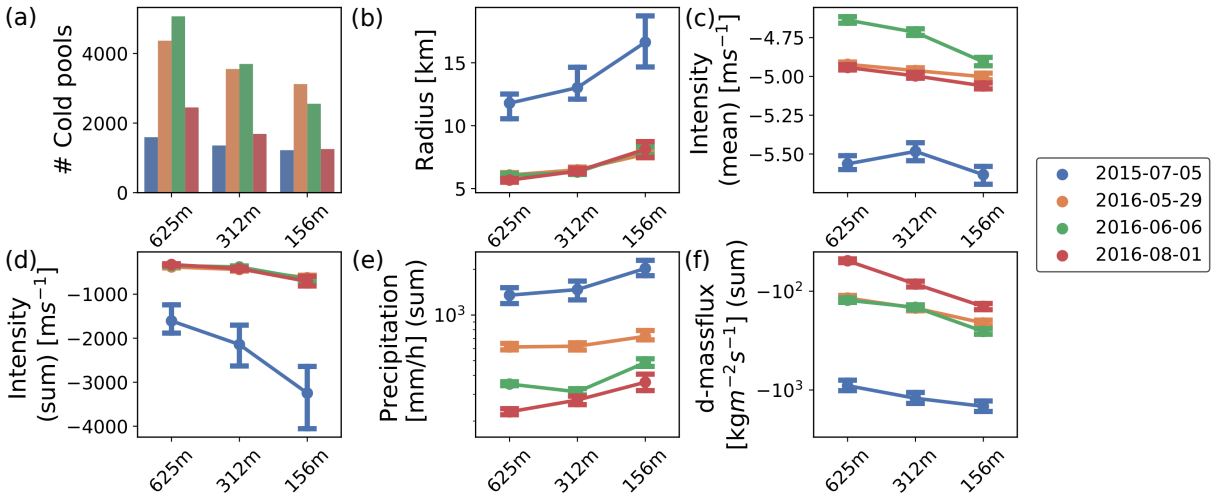


**Figure 32:** Overview of differences in cold pool populations: a) number of cold pools; b) equivalent cold pool radius (derived from cold pool area); c) cold pool mean intensity; d) cold pool integrated intensity; e) cold pool integrated precipitation and f) cold pool integrated downward mass flux. Except for the cold pool frequency, the median of the distributions is displayed with 95 % confidence intervals estimated by bootstrapping.

Having seen systematic dependencies of cold pool properties on the resolution, we now look for corresponding sensitivities in convective initiation and cold pool edge properties that may impact initiation. Comparing the model resolutions in Fig. 31b the last two days show a substantial sensitivity to the relative importance of cold pool driven convective initiation in the cold pool boundary region. This sensitivity becomes especially evident in the evening, when the higher resolutions are associated with more persistent convective activity (6 June 2016), or simply a higher peak (1 August 2016). Again we consider cold pool objects and focus now on the characteristics of their edges. Medians of convective initiation related variables are displayed in Fig 33, namely triggering probability, mean and total upward mass flux, and 95th percentiles of vertical velocity and buoyancy. Fig 33 shows a modest tendency towards increasing triggering probability with higher model resolution. Mean and total upward mass flux and 95th percentile of vertical velocity increase systematically with the resolution, although the strength of the increase varies between days. In contrast, the 95th percentile of buoyancy at the cold pool edge does not show a systematic dependence. While the first day shows enhanced buoyancy in higher resolutions, the buoyancy on the other three days is relatively insensitive, or even decreasing slightly with model resolution.

Concerning our second research question, we find that coarser models have more, but smaller and less intense cold pools. Their cold pools also have weaker gust fronts (with respect to upward mass flux/vertical velocity) and a tendency towards reduced initiation of convection. The combination of these differences are also evident in diurnal cycles of total cold pool precipitation.

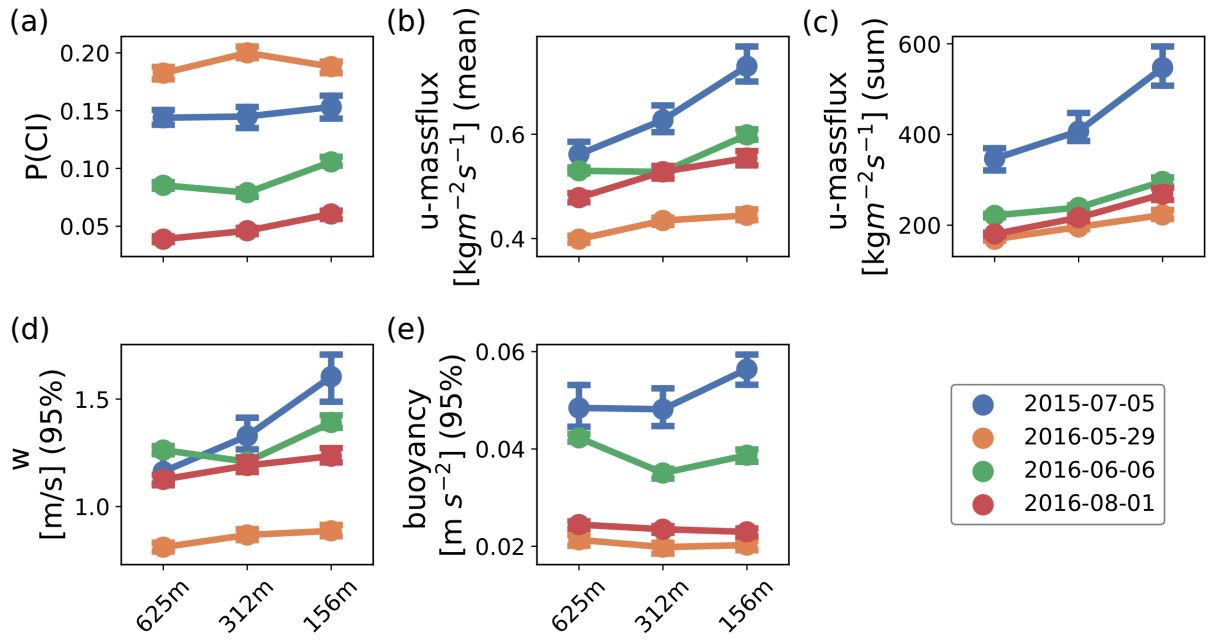


**Figure 33:** Overview of differences in cold pool populations with regard to their boundary regions and convective initiation: a) triggering probability; b) upward mass flux averaged over each cold pool boundary (1 km); c) upward mass flux integrated of the cold pool boundary area; d) 95th percentile vertical velocity (1 km) at cold pool boundaries; e) 95th percentile of buoyancy (150 m) at cold pool boundaries. Median of the distributions is displayed with 95 % confidence intervals estimated by bootstrapping.

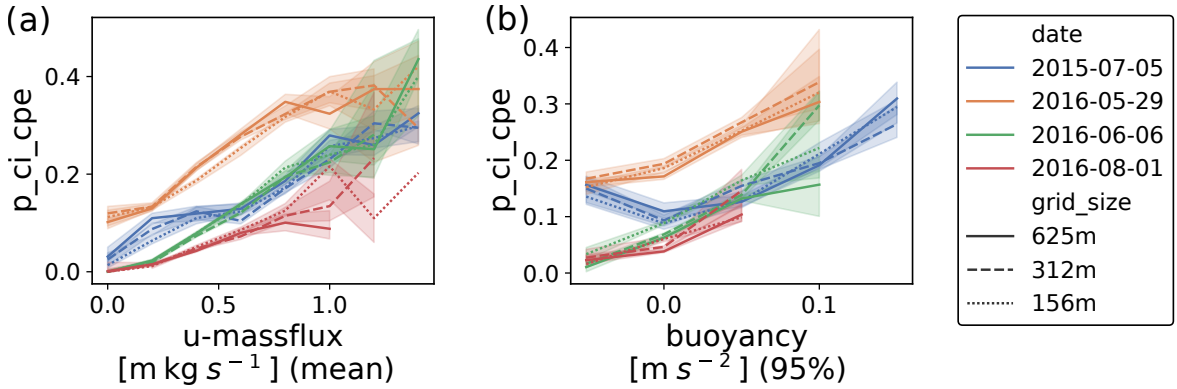


**Figure 34:** Dependence of triggering probability at cold pool boundaries to a) mean upward mass flux (1 km) and b) 95th percentile buoyancy (150 m) at cold pool boundaries. Triggering probability has been binned according to mass flux and buoyancy. To reduce sparse data ranges, values below or above the displayed ranges of mass flux and buoyancy have been set to the threshold values.

### 5.2.3 *Identifying causes of the resolution dependence of convective initiation*

Given the results from the previous section, can we conclude that the weaker vertical velocity in coarser models causes a reduction of convective initiation? Fig. 34a shows that higher mean upward mass flux is related to higher triggering, which would endorse such a conclusion. However, the triggering differences might be caused by differences in cold pool intensities. And while we observe differences between gust front strengths for all days, the difference in convective initiation is only clear for the last two days. Fig. 34b further seems to suggest that an increase in buoyancy is related to more convective initiation. Can we still neglect buoyancy because differences between resolutions are not as systematic?

To disentangle such different aspects, we follow the approach of Pearl (2013); Chen and Pearl (2015) to investigate the causal pathways under the assumption of linearity (as described in Sec. 5.1). Our goal is to identify how model resolution $R$ impacts convective initiation $P[CI]$. We distinguish between the formulation of our causal model (Sec. 5.2.3.1) and the estimation of causal path coefficients (Sec. 5.2.3.2). In Section 5.2.3.1, we describe the design of a causal graph representing our conceptual understanding of the relevant pathways through which changes in model resolution may impact convective initiation. We identify which indirect effects are reasonable and which confounding variables exist, and specify a causal graph structure with several acyclic, directed pathways from resolution $R$ to triggering probability $P[CI]$. In Section 5.2.3.2, we quantify these different pathways by applying multi-linear regression. After normalizing the data by subtracting the mean and dividing by standard deviation, we estimate each path coefficient using linear regression and include additional variables as predictors, as specified by the single door criterion. Aggregated pathways are then derived by multiplication or addition of sequential or parallel paths. For statistical uncertainty quantification, we apply bootstrapping (N = 1000). The magnitudes of the different, aggregated pathways allow us to draw conclusions about the relative importance of each pathway. To evaluate the robustness of our causal model and the subsequent estimated causal effects, we further evaluate the sensitivity of our model to some modifications (Sec. 5.2.3.3).

#### 5.2.3.1 *Causal structure*

To determine the dominant effect of model resolution on triggering probability by cold pools, we propose a causal structure, as illustrated in Fig. 35. The selected variables of interest are the model resolution (R), cold pool integrated intensity (I), cold pool boundary 95th percentile buoyancy (B), mean upward mass flux at cold pool boundaries (G), and triggering probability (P) at cold pool boundaries.

MODEL RESOLUTION Our goal is to identify the main pathway through which model resolution (R) impacts convective initiation (P). A direct effect is represented by the arrow from R to P. This pathway could also incorporate indirect pathways that are not represented in the causal model, e.g., differences in vertical acceleration, condensation processes or entrainment/detrainment due to model resolution. We have further observed that model resolution affects cold pool intensity and vertical velocity. These dependencies can be related to microphysical processes, surface

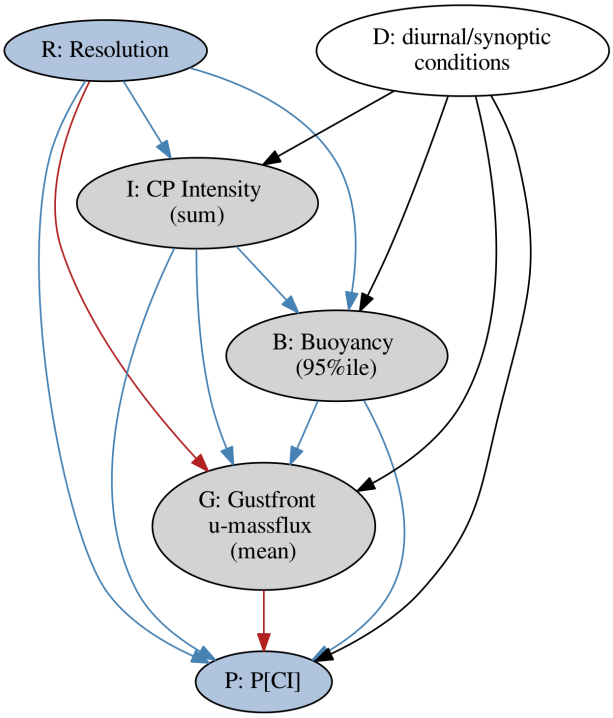


**Figure 35:** Conceptual causal model of how model resolution can impact convective initiation. Black arrows are used to emphasize their role as common driver. Blue/red arrows are used for paths that influence how model resolution (R) impacts convective initiation (P). The red paths display the below identified dominant pathway. See text for explanation.

fluxes or vertical acceleration. For buoyancy, the influence was not found to be very systematic, but we evaluate its strength in this model anyway. These causal relationships are represented by the arrows from R to I, G and B.

COLD POOL INTENSITY It is reasonable to assume that stronger cold pools have different buoyancy (B) or vertical velocity (G) at the cold pool boundary. Hence, we draw two paths IB and IG. We have further included a direct arrow from cold pool intensity to convective initiation to evaluate whether we are missing some important processes that are not captured by buoyancy or vertical velocity (IP).

BUOYANCY Buoyancy certainly influences vertical velocity, but there might also be a direct effect on triggering. These effects are represented by the arrows BG and BP.

GUSTFRONT The impact of gust front upward mass flux (mean) on convective initiation is represented by the arrow GP.

TRIGGERING We don't consider any impact of triggering on other variables. Hence there is no feedback, and we consequently have a directed acyclic graph. This is a necessary feature to enable the following estimation of causal pathways.

Furthermore, we have included the synoptic situation and diurnal cycle (D), as represented by the variables given in Fig. 26. This is necessary because we expect them to be common drivers for cold pool intensity, buoyancy, gust front and convective initiation. Time of day, however, is not taken into account explicitly.

Several noteworthy assumptions have been made with regard to the causal structure. First, no temporal evolution or time lag is considered. Hence we assume that the

autocorrelation time scale of the variables of interest is longer than the time it takes for stronger cold pool intensities to affect gust fronts, buoyancy or convective initiation. This also means, that the data are not independent, as subsequent time steps may include the same, temporally correlated cold pool. The causal time series approach introduced by Runge *et al.* (2015) is likely better suited for time series, and we consider its application to tracked cold pools in the future. We also neglect any feedback cycles, e.g., that stronger convective initiation might lead to a strengthening of cold pools. This is justified if the autocorrelation time scale of the variables of interest is expected to be smaller than the 15–30 min timescale for precipitation development. Furthermore, no impact of model resolution on synoptic conditions is assumed (no link between R and D). This is justified by the small resolution sensitivity of the domain aggregated, large scale variables in Figure 26. Finally, we assume that we have captured all major variables that have a strong influence on at least two of the variables of interest (assumption of no unobserved common drivers).

#### 5.2.3.2 *Estimating causal effects*

To estimate the path coefficients of the causal graph, we apply multiple linear regression based on the single door criterion using all identified cold pools at all time steps. Table 4 displays the predictor sets considered for each path. The resulting average path estimates are illustrated in Fig. 36a. As expected, strong paths are observed between buoyancy and gust front upward mass flux, and also between gust front mass flux and triggering probability. The impact of buoyancy on triggering probability, however, seems to be completely mediated by the gust front. More intense cold pools lead to an increase in buoyancy at the cold pool boundaries and a small decrease in upward mass flux (the signs of these paths arise from the fact that intensity is negative). The graph further shows that, in comparison to synoptic conditions, the model resolution is of secondary importance for the variables of interest. Hence, all models are able to capture the dominant, physical processes. This is to be expected and does not undermine the importance of comparably small, systematic biases due to model resolution.

To evaluate dominant direct and indirect pathways between resolution and triggering probability, aggregated path estimates are displayed in Fig. 36b. The $*$ in $R * B * P$ denotes that all possible direct or indirect paths between the two specified variables are accounted for. This shows that the $RP$, $RIP$, $RIGP$ and $R * B * P$ paths are relatively small. Hence, the dependence of the cold pool intensity and the buoyancy on resolution do not play an important role in triggering probability. The $RGP$ path clearly dominates: larger grid sizes decrease upward mass flux at the cold pool boundaries, which consequently reduces convective initiation. The displayed, de-normalized value of $-0.017$ means that if model grid size is increased from 156 to 625 m, triggering probability will be reduced by 0.017. This is somewhat less than some of the differences observed in Fig. 33 but still of comparable magnitude. Furthermore, this corresponds to a difference of 10-30% in median triggering probability.

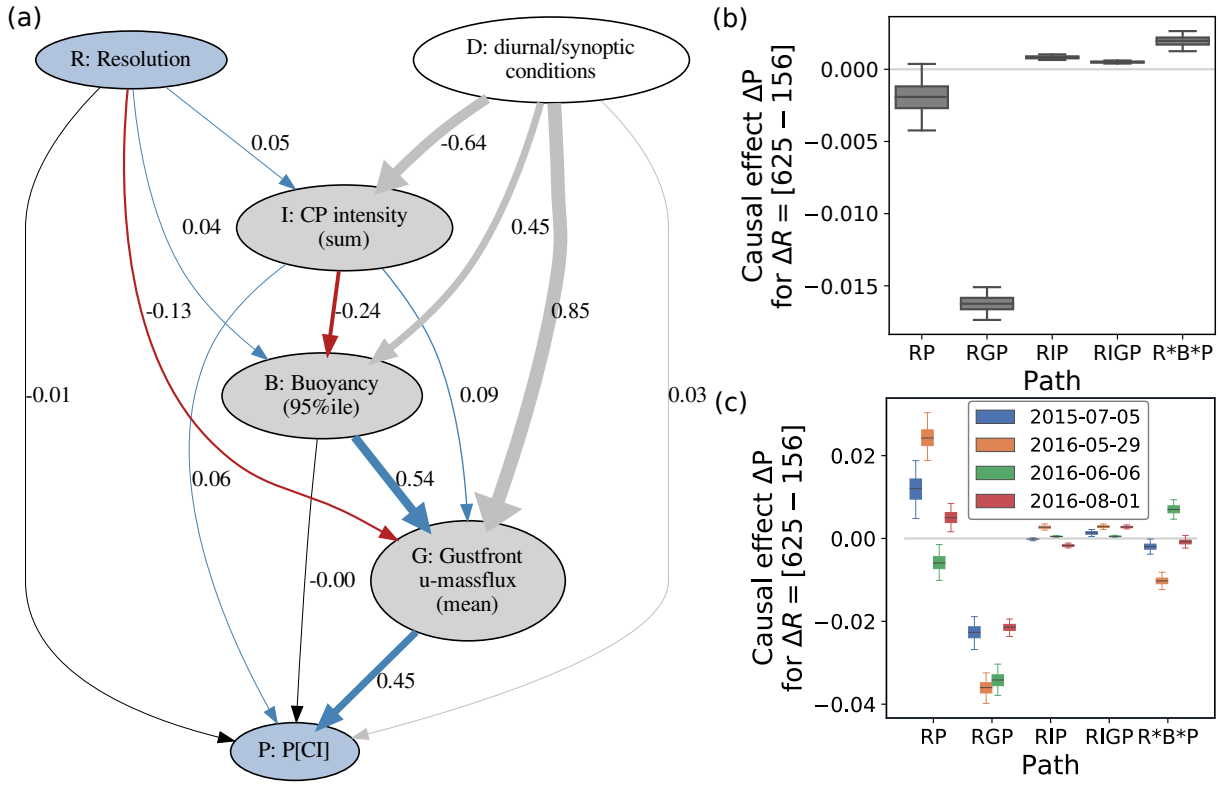


**Figure 36:** (a) Causal structural model of how model resolution can impact convective initiation. Single estimated path coefficients (normalized) are displayed and qualitatively visualized by the thickness of the respective paths. Negative arrows (red) indicate that an increase in the parent variable results in a decrease in the child variable. The paths from the diurnal/synoptic variables display the sums over all diurnal/synoptic variables for better visualization and are colored in grey. (b) Aggregated path estimates, de-normalized to represent the causal effect of changing model resolution by $625\,m - 156\,m$ on triggering probability. (c) as (b), but with stratification by days. Note that (b) can not be obtained simply as the average over the four days of (c). The whiskers of the box plots correspond to the 95 % confidence intervals from bootstrapping.

#### 5.2.3.3 *Robustness*

The statistical robustness of the results and their sensitivity to several of the assumptions made in constructing the causal graph are examined in Appendix A.2. This includes tests of the statistical significance of the linear regression results, the linearity of the response to changing model resolution, selecting different variables as proxies for gust front or cold pool intensity, sensitivity to stratification by days, and of the impact of model resolution on the synoptic conditions. One noteworthy concern arises from sometimes relatively small $R^2$ values, which can imply that several informative variables may not be included in our causal graph. This can increase the likelihood of missing common driver, which causes biased causal effect estimates (see Appendix A.2 for more details).

Nonetheless, the causal effect estimates vary moderately under most modifications with no strong dependence on the specific resolution range or the selected variables. Only the stratification of the data into the four individual days substantially influences the causal effects. The path estimates computed for each of the four days are illustrated in Fig 36c. The estimates for the paths *RIP* and *RIGP* show some variability between days, but the values are all close to zero. The path estimates of the *R*B*P* path show stronger variability among the days, although centered around zero. The *RGP* path, which we have earlier identified as the dominant path, shows moderate variability between the days, with values ranging from -0.02 to -0.04. Thus, this effect is relatively robust to the stratification by days and is consistently stronger than other indirect effects. On the other hand, the direct effect of resolution on triggering (*RP*) shows a strong sensitivity to the day, ranging from 0.03 to -0.01. Since the direct *RP* path includes influences of causal processes not represented explicitly in the graph, it is interesting to note that it is strongest on 29 May 2016, when the diurnal cycle of precipitation is disturbed by strong synoptic forcing. In such situations, the initiation of convection may be less influenced by cold pool effects.

Overall, we conclude that the relatively robust *RGP* path predominates in most situations. Hence, to answer our third research question, the reduction in cold pool driven convective initiation in coarser models is a direct result of reduced upward mass flux at cold pool boundaries. The eventual impact on precipitation, however, is very dependent on the weather situation, since the dominant *RGP* path can be partially compensated by other processes.

## 5.3 SUMMARY AND DISCUSSION

This work aims to identify the relevance of cold pools for the diurnal cycle and the organization of convection; whether convection-permitting models can adequately represent cold pool driven convective initiation; and which misrepresented processes dominate the deficiencies in cold pool driven convective initiation. To do so, we compared available high-resolution ICON simulations with grid sizes 156 m, 312 m and 625 m of four selected days over Germany and identified cold pools, cold pool boundaries and convective initiation locations.

The first research question was to determine whether cold pools are a significant factor for the diurnal cycle and organization of convection. To address this, we have evaluated identified cold pools and quantified the related precipitation and convective initiation. We found strong links between precipitation and cold pools. Although day-to-day variability is high, cold pool driven convective initiation can be as much as 50% of the total convective initiation. For two of the four days, this is especially prominent in the late afternoon and evening. Cold pool boundary regions, where gust fronts occur, were identified as very efficient convection initiators, with triggering probabilities several times higher than outside cold pools. Consistent with previous studies (e.g., Böing *et al.*, 2012; Schlemmer and Hohenegger, 2014; Feng *et al.*, 2015), these results demonstrate the importance of cold pools for convective organization and the diurnal cycle of convection, but also the strong role of the synoptic environment.

The second research question asks whether convection-permitting models can accurately represent the convective initiation process. This was addressed by evaluating the sensitivity of cold pool related properties to model resolution, using simulations with grid sizes from 156 m, where there is some hope of resolving the relevant processes, to 625 m, where the characteristic errors of convection-permitting models should be apparent. For two of the four days we identified a significant sensitivity of the diurnal cycle of precipitation and convective initiation to model resolution. The cold pools themselves were found to be smaller and less intense at coarser resolutions, which is in agreement with the results from Squitieri and Gallus (2020). Regarding the influence of model resolution on convective initiation, a reduced triggering probability with increasing grid size was found for two of the four simulated days. Vertical velocity and upward mass flux at the gust fronts were reduced with increasing grid sizes, while buoyancy in this region did not show a systematic sensitivity. These findings show that coarser models have difficulties in representing cold pool driven convective initiation. The specific character, however, seems to depend on the large scale situation.

The third research question is concerned with which errors in the coarse resolution simulations lead to the observed differences in convective initiation with changing model resolution. To evaluate the relative importance of different causal mechanisms, we applied a linear causal effect estimation based on causal graphs. Again the day-to-day variability is large, but a systematic and dominant causal effect was identified: Coarser models yield weaker updrafts at the cold pool gust fronts, which then cause reduced convective initiation. Since the intensity of the cold pools and the buoyancy of the lifted air do not appear to play a role in the sensitivity, it is likely that the reduced upward motion at the boundary is simply a result of insufficiently resolving sharp gradients at the gust front.

The causal analysis is based on several assumptions of the underlying model, that have been mentioned and tested to some extent in Section 5.2.3.1, 5.2.3.3 and Appendix A.2. Open issues include the possibility of missing common drivers, e.g., due to neglecting the temporal evolution of cold pools or due to missing other variables. This possibility is strengthened by moderately high $R^2$ values (see Appendix A.2). Furthermore, by using identified cold pools from 5 min model output, subsequent time steps include the same cold pools, and the data is not independent. This may cause an overestimation of the significance. Extensive testing of the robustness of the causal analysis

revealed one important sensitivity, namely the dependence on the different synoptic conditions present on the four days. This resulted in a large variability of the direct effect of model resolution on triggering probability (RP). On some days, this direct effect partially compensates the effect of reduced convective initiation by weaker gust fronts (RGP), whereas on other days, the effect is negligible or complements the RGP path. Since the direct path includes the effects of mechanisms not represented in the causal graph (which was focused on cold pool processes), it may become important when strong synoptic forcing leads to convective quasi-equilibrium, and local triggering processes are less relevant (Keil *et al.*, 2014). To investigate such processes thoroughly will require a much larger number of simulation days than are currently available. It is worth recalling, however, that there was no day in the current analysis where the RP path became as strong as the dominant RGP path.

While we investigated the causes of the reduced triggering probability for our third research question, we have not addressed the causes of the differences in cold pool sizes and intensities. We speculate that the following processes may be relevant. First, the representation of cloud microphysics can have a substantial influence on precipitating downdrafts or on sub-cloud layer evaporation, which can, in turn, influence cold pool numbers, sizes and intensities. Entrainment or detrainment into or from the downdraft or the cold pool can also be relevant. They are mostly determined by turbulence parameterizations, which show substantial deficiencies even in models with resolutions of hundreds of meters, often denoted as the terra incognita, or grey zone, of turbulence parameterization (Wyngaard, 2004; Honnert, 2016). Surface fluxes have also been found to have a significant impact on cold pools (Gentine *et al.*, 2018; Grant and van den Heever, 2018) and can be sensitive to the resolution of small scale surface features. Finally, stronger convective initiation at the gust front can result in reinforcement of the original cold pool and thereby increase cold pool size and intensity (Böing *et al.*, 2012; Schlemmer and Hohenegger, 2014). Further analysis of the statistics of cold pools and the mechanisms that influence them within the ICON simulations is planned for a future publication.

The original motivation for this study came from systematic errors observed in kilometer-scale numerical weather prediction models. The extrapolation of the present results to such models is based on two critical assumptions. First, the highest resolution simulation examined here must behave sufficiently similar to the real atmosphere to serve as a reference. A thorough evaluation of the ICON configuration used here (Heinze *et al.*, 2017) confirms that the highest resolution configuration is able to reproduce observed turbulence profiles well and generates cloud size distributions that are more realistic than for the coarser resolutions. Pscheidt *et al.* (2019) still find some deficits in precipitation and cloud coverage, but convective organization seems to be well captured. Second, the errors in the coarser-resolution runs must be representative of errors found in the NWP models, if the results are to provide guidance for correcting those errors. One concern regards the different approaches for turbulence parameterizations. While a three-dimensional, LES-type turbulence closure (Heinze *et al.*, 2017) was employed in the hectometer-scale simulations analysed, most km-scale models, including the COSMO model (Chapter 2), are equipped with a one-dimensional PBL-scheme (e.g., Kealy, 2019, or Chapter 1.3.2). However, both hectometer- and km-scale models

suffer from similar deficiencies: the lag in timing of the diurnal cycle of convection and a relative lack of convective organization, especially late in the day (e.g., Baldauf *et al.*, 2011; Clark *et al.*, 2016; Rasp *et al.*, 2018b). Hence we conclude, that the here obtained biases in cold pool driven convective initiation for hectometer-simulations are valid and likely even more pronounced in km-scale models and that they are partially responsible for biases in precipitation forecasts. However, given the large day to day variability seen here, future studies using much longer time periods are desirable.

In the next chapter, we will now use the gained understanding to improve the representation of cold pool driven convection in the km-scale model COSMO. We expect that the observed systematic errors can thereby be reduced.

# 6

# COLD POOL DRIVEN CONVECTIVE INITIATION (II): REDUCING MODEL DEFICITS

Now we turn again to km-scale models and address how we can improve the inadequate cold pool driven convective initiation in such models.

As already described in Chapter 1.3.4 and confirmed in chapter 5, cold pool gust fronts provide efficient triggers for new convection in the vicinity of already existing convection. As a consequence, cold pools contribute to the organization of convection and provide enhanced trigger mechanisms in the late afternoon, which impacts the diurnal cycle of convection.

Both a lack of organization and difficulties in capturing the diurnal cycle have been identified in km-scale models by several studies. Missing convective organization, structural deficits in cloud sizes or biases in the onset of convection and the afternoon and evening amplitude were found (Baldauf *et al.*, 2011; Hanley *et al.*, 2015; Clark *et al.*, 2016; Senf *et al.*, 2018; Rasp *et al.*, 2018b; Panosetti *et al.*, 2019).

That cold pools will be responsible for at least part of these deficits was further supported by results from the previous chapter. In the previous chapter, we evaluated the sensitivity of cold pools and cold pool driven convective initiation to model resolution. We found that cold pools are more frequent, but smaller, less intense and trigger less new convection in coarser resolutions. Specifically, we found that the dominant deficit in triggering probabilities comes from too weak gust fronts. Total cold pool driven convective initiation is likely further affected by the size of the cold pools. Smaller cold pools mean less total convective initiation even if the gust fronts have the same triggering probabilities.

Building on these previous results, we now aim to improve the representation of cold pools in km-scale models. Few such cold pool parameterizations already exist for convection parameterized models, but their application in convection-permitting simulations is generally not valid (Rozbicki *et al.*, 1999; Grandpeix and Lafore, 2010; Grandpeix *et al.*, 2010; Park, 2014). The ParaCon project (ParaCon, 2020) is currently developing a new type of scale-adaptive convection parameterizations for the convective grey zone, which will include parameterized cold pools to enhance convective initiation (personal communication). Although they pursue a scale adaptive grey zone parameterization, their parameterized cold pools will likely coincide with crudely resolved cold pools. Here, we explicitly target km-scale models where cold pools are comparably well resolved, and - instead of parameterizing whole cold pools - we aim to modify resolved cold pools in a more subtle way to obtain more realistic triggering of convection at their gust fronts.

Based on the results from the previous chapter (Chapter 5) that differences in cold pool intensity only marginally impact the triggering probability (RIGP path) whereas the resolution directly affects gust front strength and thereby triggering (RGP path)

(see Fig. 36), we apply the following strategy to improve cold pool driven convective initiation: by perturbing vertical velocity tendencies at gust front regions in the model, the gust fronts will be strengthened. We expect this strengthening to directly enhance convective initiation at gust front regions. Additional, indirect benefit may occur as stronger cold pool driven convective initiation may further strengthen the original convective system and thereby also enhance cold pool size and intensity. This may further improve convective initiation, resulting in a feedback loop (Böing *et al.*, 2012; Schlemmer and Hohenegger, 2014).

We summarize the primary research goals of this chapter as follows:

1. To develop a cold pool perturbation scheme, the CPP scheme, to improve cold pool driven convective initiation.

2. To identify the impact of CPP on different aspects of the model, most importantly, the afternoon/evening precipitation and its organization.

To address these research goals, we will test the CPP scheme on one single day with several different parameter settings and a different model configuration. For a reduced selection of settings, we apply the CPP scheme also for the ten days considered in chapters 2 and 3. We identify how CPP affects the diurnal cycle and organization of convection and whether it reduces the mentioned biases. We will further evaluate the combination of the CPP scheme with the PSP2, as a complementary process representation (subgrid-scale variability of boundary-layer turbulence) to CPP (see Chapter 3). Hence their combined impact is expected to yield additional benefit compared to using either CPP or PSP2.

This chapter is structured as follows. First, we give a short overview of the selected days and the simulation strategy in Section 6.1. Next, we will describe some theoretical considerations on the resolution dependence of cold pool gust fronts in Section 6.2.1, which we then use to formulate the CPP scheme in Section 6.2. Then, we will evaluate the impact of the CPP scheme in Section 6.3 and end with a discussion in Section 6.4.

## 6.1 STRATEGY FOR DEVELOPING CPP AND EVALUATING ITS IMPACT

We develop CPP and evaluate its impact first for one single day, 5 June 2016. This day represents a day with many cold pools, and hence the impact of CPP is expected to be strong. We further use this day to investigate the sensitivity of the CPP impact to changes in some of its parameters and model settings. Based on these simulations, we then select reasonable parameters to test CPP for ten subsequent days, each with 24 h simulation time to confirm our findings and evaluate the flow dependency of CPP. As for Chapter 3 and 4, we will use the COSMO model over Germany and compare the precipitation fields with radar observations. Aside from domain averaged precipitation fields, we consider the fraction skill score (FSS) as a diagnostic for the spatial location of precipitation, the structure component of the SAL score (S-SAL), average cell sizes, and frequency to identify changes in the diurnal cycle and organization of convection. Further details on the selected days, the COSMO model, radar data and diagnostics are given in Chapter 2.

## 6.2 COLD POOL PERTURBATIONS CPP

In the previous chapter we have seen that lower resolution simulations yield weaker vertical velocities at cold pool gust fronts. Hence, their gust fronts are expected to be underresolved. As correcting measure, the cold pool perturbations aim to impose vertical velocity perturbations to achieve the following: lower resolution simulations of otherwise underresolved gust fronts will then produce vertical velocity scales of fully resolved cold pool gust fronts. To get a better understanding of how grid size affects the vertical velocity scale of gust fronts we use dimensional analysis in Section 6.2.1. This also provides a target vertical velocity scale $w_0$ of fully resolved cold pools. In the subsequent subsections, we use this target vertical velocity scale $w_0$ to build appropriate cold pool perturbations for the numerical model.

### 6.2.1 *Scale of gust front vertical velocity*

For the dimensional analysis, the inviscid, Boussinesque approximated set of equations for the stream function $\Phi$, vorticity $\eta$ and vertical velocity $w$ in a two dimensional plane in x and z are used, which are often considered for density currents (Rotunno *et al.*, 1988; Weisman and Rotunno, 2004; Bryan and Rotunno, 2014b). Characteristic scales are used to non-dimensionalize the equations. A detailed derivation is given in Appendix B. The following characteristic scale $W$ for the vertical velocity is then given as a function of the characteristic buoyancy scale $B$, the horizontal length scale $L$ and the vertical length scale $H$:

$$W = \sqrt{\frac{BH}{1 + \frac{L^2}{H^2}}}. \tag{11}$$

This relationship describes the characteristic, vertical velocity scale within a circulation driven by horizontal buoyancy gradients and how it depends on both the horizontal and vertical length scales in addition to the characteristic buoyancy scale.

Similar relationships have been formulated for rising warm bubbles by Weisman *et al.* (1997); Pauluis and Garner (2006); Morrison (2016a,b); Jeevanjee and Romps (2016); Jeevanjee (2017). These studies investigate buoyant acceleration and associated pressure perturbations, which arise from displacing the air above the vertically accelerating parcel to make way for it. These pressure perturbations can reduce the effective acceleration substantially. Importantly, these pressure perturbations depend on the horizontal extent L and the vertical extent H of the buoyancy acceleration: for thin and tall parcels ($\frac{L}{H} \approx 1$), the pressure perturbation deceleration is moderate while for wide and flat parcels ($\frac{L}{H} \gg 1$) the pressure perturbation deceleration is larger as substantially more air has to be displaced. Dependent on the specific formulation of the problem, the arising dependency of $W$ on $\frac{L}{H}$ varies in the different studies. However, Jeevanjee (2017) showed that the different scalings from Weisman *et al.* (1997); Pauluis and Garner (2006); Morrison (2016a); Jeevanjee (2017) are comparable. We have further confirmed a close resemblance of (11) with the formulations of the mentioned studies. As the underlying physical processes for the generation of vertical motion are comparable for cold pool gust fronts, this resemblance is not surprising.

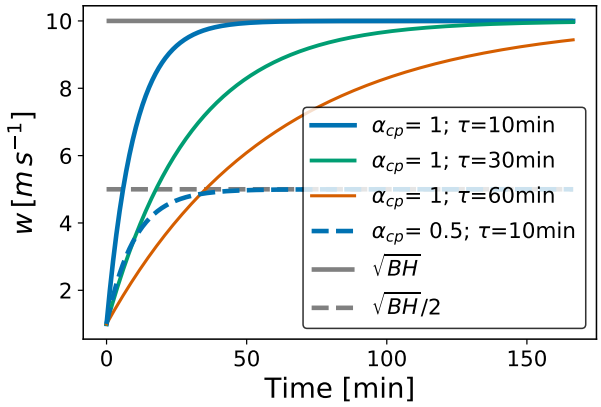


**Figure 37:** Schematic illustration of the temporal evolution of $w$ and how it depends on $\tau_{cp}$ and $\alpha_{cp}$. Here, we consider the isolated situation of $\frac{\partial w}{\partial t}\Big|_{cp}$ when no other effects are included, and $w_0$ remains constant.

Within current NWP models, the horizontal scale $L$ is limited by the horizontal grid size $\Delta x$, explaining resolution sensitivities and weaker vertical velocities with coarser resolutions.

In the case of a fully resolved gust front, we assume $\frac{L}{H} \approx 1$. Then we obtain a characteristic vertical velocity scale $w_0$:

$$w_0 = \sqrt{\frac{BH}{2}}.$$

This relationship is further comparable to the well-established propagation speed of density currents $U = \sqrt{2BH}$ (e.g., von Kármán, 1940; Benjamin, 1968; Bryan and Rotunno, 2008; Markowski and Richardson, 2011, chap. 5.3.2).

### 6.2.2 *Basic approach of CPP*

We now develop a pragmatic approach, the CPP, to remedy the deficit of $W$ in coarser model resolutions. To do so, we apply tendency perturbations $\frac{\partial w}{\partial t}\Big|_{cp}$ to the model vertical velocity $w$ so that the amplitude of $w$ will converge to the target $w_0$ on a time scale determined by $\tau_{cp}$:

$$\frac{\partial w}{\partial t}\bigg|_{cp} = \frac{1}{\tau_{cp}}(w_0 - w).$$

This behavior is schematically illustrated in Fig. 37 for the isolated situation where all other effects are neglected.[12] To ensure a persistent impact on $w$ as for the PSP2 and SSOSP scheme, we further perturb tendencies of the horizontal wind fields $u$ and $v$ to yield three dimensional, non-divergent perturbations. Details can be found in Sec. 3.1.2.4.

Given this basic framework, the following details have to be considered, which will be addressed separately in the next subsections:

1. Approximating and tuning $w_0$

12 A similar perturbation structure is sometimes used to nudge the model towards observations, e.g., latent heat nudging (see, e.g., Schraff *et al.*, 2016), and sometimes referred to as Newtonian relaxation.

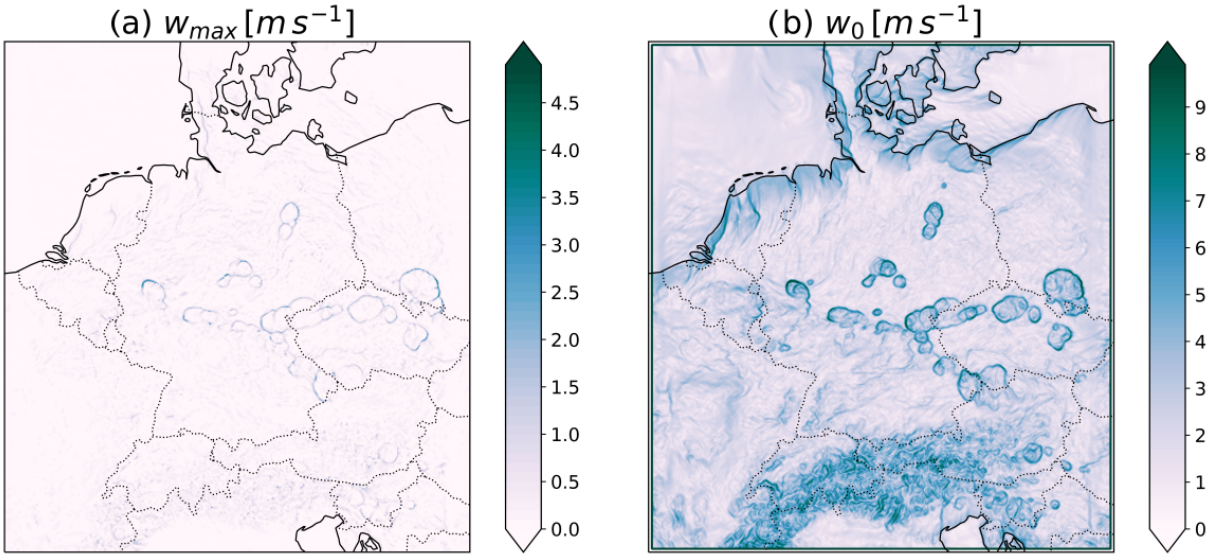


**Figure 38:** Comparison of maximum vertical velocity $w_{max}$ in the lowest kilometer at the surface (a) and the target vertical velocity $w_0$ ($\alpha_{cp} = 1$) as approximated from $\theta_v$ gradients (b). Note the different scales of the colorbars.

2. Considering the vertical distribution of the perturbations
3. Finding a reasonable timescale $\tau_{cp}$
4. Confining perturbations to cold pool gust fronts

### 6.2.3 *Approximation of $w_0$ and $\alpha_{cp}$*

To compute the target vertical velocity for CPP, $w_0 = \sqrt{\frac{BH}{2}}$, we have to approximate the cold pool buoyancy $B$ and its height $H$. For simplicity, we fix $H = 200\,m$. This height corresponds approximately to the buoyancy anomaly of cold pools within the COSMO simulations (not shown) and has also been used in the ICON simulations in Chapter 5. As we do not intend to identify cold pools online in the model, we estimate the cold pool buoyancy as the local buoyancy gradient near the surface multiplied by the width of the gradient. As the model cannot resolve sharp gradients, we assume that gradients will be flattened on a scale of the effective model resolution of $5\Delta x$. We define $B$ as:

$$B = |\nabla\overline{\theta_v}| \cdot 5\Delta x g \frac{1}{\overline{\theta_v}},$$

where the bar denotes the mass-weighted average over the lowest 5 model levels, which corresponds approximately to the lowest 200 m.[13]

Doing so, we obtain $\sqrt{BH}$ with values of up to $\mathcal{O}(10\,m\,s^{-1})$ at cold pool gust fronts, as displayed in Fig. 38b. While this is significantly stronger than what is currently simulated for $w$ with the model (Fig. 38a), and even with the 156 m ICON-LEM simulation

13 Instead of using the environmental $\theta_v$ as reference temperature (as is usually considered for buoyancy), we use the local values instead. With temperature differences of few Kelvin compared to reference temperatures of $\mathcal{O}(300K)$, this is a reasonable assumption and strongly simplifies the computation.

from Chapter 5, it provides a characteristic scale, that can be regarded as the upper limit. Aspects, such as friction or entrainment, which expectedly reduce the vertical velocity, have not been taken into account. Also, several strong approximations have led to the computation of $\sqrt{BH}$. Hence, we allow for some flexibility with regard to $w_0$ by introducing the tuning parameter $\alpha_{cp}$ as:

$$w_0 = \alpha_{cp}\sqrt{BH} \tag{12}$$

If the construction of CPP is physically reasonable, $\alpha_{cp}$ should be of magnitude $\mathcal{O}(1)$ and should generalize to other situations and model setups.

### 6.2.4 *Vertical distribution of the perturbations*

As the scale $w_0$ does not include any information on the vertical distribution of $w$, we will use the vertical distribution from the model vertical velocity itself to scale $\left.\frac{\partial w}{\partial t}\right|_{cp}$.[14] To do so, we compare $w_0$ to the maximum vertical velocity $w_{max}$ from the surface to a fixed model level $k_{ref} = 38$. This height corresponds approximately to $1070\,m$ and has been chosen as an approximate upper limit for the cold pool gust front. We assume that the difference between $w_0$ and $w_{max}$ applies to the other model levels in a multiplicative way. Fig. 39 schematically illustrates this: the relative difference $\frac{w_0 - w_{max}}{w_{max}}$ is used as multiplication factor for all other model levels. Then, we can write the CPP scheme as:

$$\left.\frac{\partial w}{\partial t}\right|_{cp} = \frac{1}{\tau}\frac{(\alpha_{cp}\sqrt{BH} - w_{max})}{w_{max}} \cdot w(z).$$

We further aim to constrain the perturbations to the layer where the cold pool gust front is active, i.e., the lowest model levels up to level $k_{ref}$. Above this model level, the perturbations are linearly tapered to zero over $\Delta H = 500\,m$, as visualized by the dark blue line in Fig. 39.

### 6.2.5 *Time scale $\tau_{cp}$*

The time scale $\tau_{cp}$ at which the target $w_0$ is approached, has to be determined under the following two constraints. First, the time scale should not be too short, as this will mean virtually instantaneous perturbations, which likely cause instabilities. In order to allow for sufficient adaptation of the boundary layer to the perturbations, the time scale should not be shorter than the lifetime of the turbulent eddies, i.e., $10\,min$. Second, the time scale should be shorter than the lifetime of the cold pool gust fronts. Otherwise, the target $w_0$ will not be realized. The lifetime of cold pool gust fronts is, however, highly variable and can be as short as several minutes and as long as few hours (e.g., Feng *et al.*, 2015). From this perspective, the time scale should be as short as possible. Consequently we consider a time scale of $\tau_{cp} = 10\,min$, but we will also evaluate the behavior of longer time scales (20 and $30\,min$).

14 For the dimensional analysis we can formulate the dimensional vertical velocity $w$ as $w = W(B, H, L)\tilde{w}$, where $\tilde{w}$ represents a non-dimensional function, that only depends on other non-dimensional variables. This justifies the use of the vertical profile for $w$ from the model also for the perturbations.

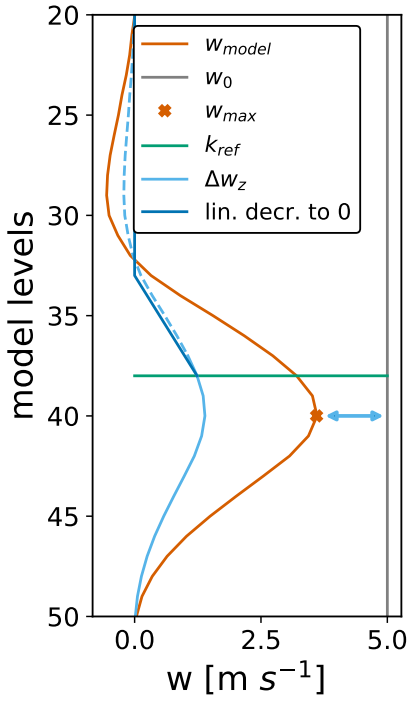


**Figure 39:** Schematic illustration of how the vertical structure of the perturbations is designed. $w_{model}$ (blue line) represents a possible vertical structure of vertical velocity at a cold pool gust front. Its difference to $w_0$ at a given height $k(w_{max})$ is used as the target $\Delta w$ for this height. $\Delta w_z$ for all other heights are then scaled according to the $w_{model}$ profile. Above a certain height ($k_{ref} = 38$ corresponds to $\approx 1\,km$), the perturbations are linearly decreased to zero (dark blue line).

### 6.2.6 *Limiting perturbations to cold pools*

Now, we describe how we constrain the perturbations to cold pool gust fronts. Without having to identify cold pools or cold pool gust fronts directly, we apply the following criteria.

First, to avoid perturbing areas with very weak $\theta_v$ gradients - which likely do not coincide with cold pool gust fronts and, hence, our reasoning is not valid there - we constrain the perturbations to be active only if a certain threshold of $|\nabla\theta_v|$ is exceeded, namely $\theta^*_{v,g} = 0.75\,K\,\Delta x^{-1}$. To smooth the resulting field, we apply a uniform filtering with $n_{filter} = 3$ grid boxes filter size.

Second, we identify gust fronts as grid points with strong vertical velocities and hence we require $w_{max}$ to be larger then $w^*_{max} = 0.5\,m\,s^{-1}$ for active perturbations. Example fields of these two criteria are shown in Fig. 40a, b and confirm that these criteria work well to identify cold pool gust fronts. However, we still find many active grid points over orographic regions, which are mostly not associated with cold pools.

The third criterion excludes these orographic regions by switching off the perturbations at grid points with a standard deviation of subgrid-scale orographic height $\mu_{sso}$ above a threshold $\mu^*_{sso} = 50\,m$ (see Fig. 22a for the $\mu_{sso}$ field).

We include one last constraint for grid columns, where the model vertical velocity $w_{max}$ is already larger than the target vertical velocity $w_0$. Then, the perturbations are set to zero.

The arising horizontal mask for the first three criteria is displayed in Fig. 40c. Most perturbations are now indeed related to cold pool gust fronts. We note that also land-sea-breezes are partially included, which will be discussed in more detail in Sec. 6.4. The final perturbations are displayed in Fig. 40d.

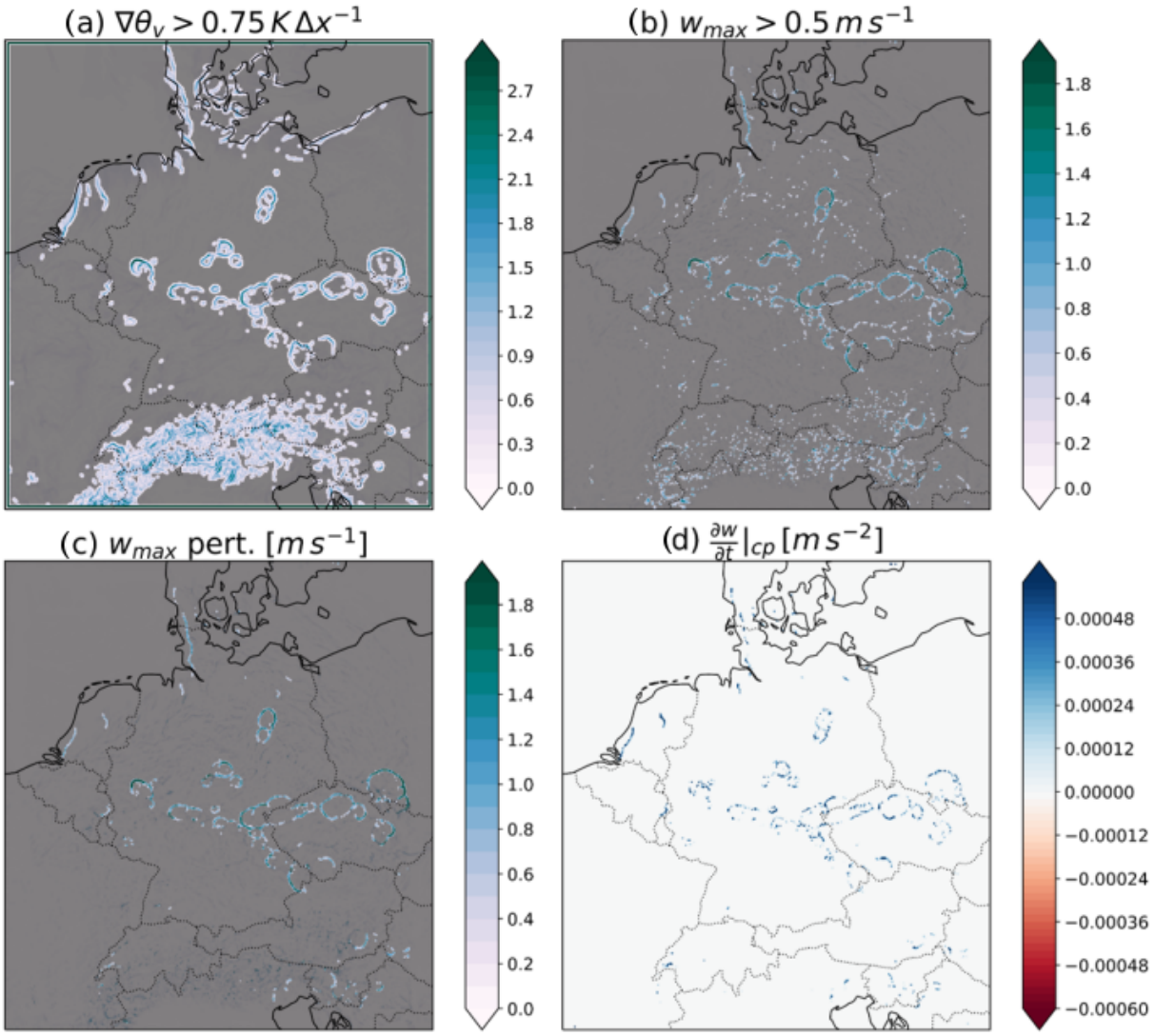


**Figure 40:** Illustration of how the perturbations are constrained to cold pool gust fronts. In (a), the buoyancy gradients are displayed with values smaller than 0.75 $K\,\Delta x^{-1}$ shaded grey. In (b), $w_{max}$ is shown with values smaller 0.5 $m\,s^{-1}$ shaded grey. In (c), $w_{max}$ is shaded using the first two criteria combined with the orography criterion. In (d), the final perturbations for $\frac{\partial w}{\partial t}$ are shown.

### 6.2.7 *Overview of parameters*

As has been explained above, several parameters have to be specified, which are listed in Table 3. As there is some uncertainty in these parameters, we will test in Sec. 6.3.4 how sensitive the CPP impact is to changes in some parameters. The selected values are further displayed in Table 3.

**Table 3:** Parameters of CPP are listed. Default values are shown in bold font. Note that we only evaluate the sensitvity of CPP to the first four parameters.

| Parameter | Description | Values | Sensitivity |
|---|---|---|---|
| $\alpha_{cp}$ | Tuning parameter to scale the target $w_0$ | 0.15, **0.2**, 0.3 | Increased values result in higher target $w_0$ and in a stronger precipitation impact. |
| $\tau_{cp}$ | Time scale over which to reach target $w_0$ | **10**, 20, 30 [min] | Increased values can be compensated by reductions in $\alpha_{tuning}$ to some extent, and vice versa. |
| $w^*_{max}$ | $w_{max} > w^*_{max}$ for perturbations to be active. | 0.3, **0.5**, 0.7 $[m\,s^{-1}]$ | Reducing $w^*_{max}$ increases horizontal area of perturbations and precipitation impact. |
| $k_{ref}$ | Maximum height level for searching $w_{max}$ and the perturbations. | **38** $\approx$ **1** $km$, 36, 34 | Reducing $k_{ref}$ (increasing the height) strengthens the impact of CPP. |
| $n_{filter}$ | Filter size for smoothing gust fronts | **3** $\Delta x$ | |
| $\theta^*_{v,g}$ | Threshold for $\theta_v$ gradient | **0.75** $K\Delta x^{-1}$ | |
| $\mu^*_{sso}$ | Threshold for extracting only non-orographic grid points | **50** m | |
| $\Delta H$ | linear decrease above $k_{ref}$ | **500** m | |

## 6.3 IMPACT OF CPP

We now first consider 5 June 2016 with default settings for CPP (see table 3) to briefly evaluate the general impact of CPP.

### 6.3.1 *Cold pool gust fronts*

The most direct effect of CPP should certainly be visible in the cold pool gust fronts. Fig. 41 shows vertical velocity five model levels ($\approx 200\,m$) above the surface. In comparison to the reference run, the gust fronts, i.e., the line structures of positive vertical velocity, are enhanced with higher vertical velocities in the CPP simulation. This snapshot indicates that the intention of the CPP - namely to strengthen cold pool gust fronts - is achieved.

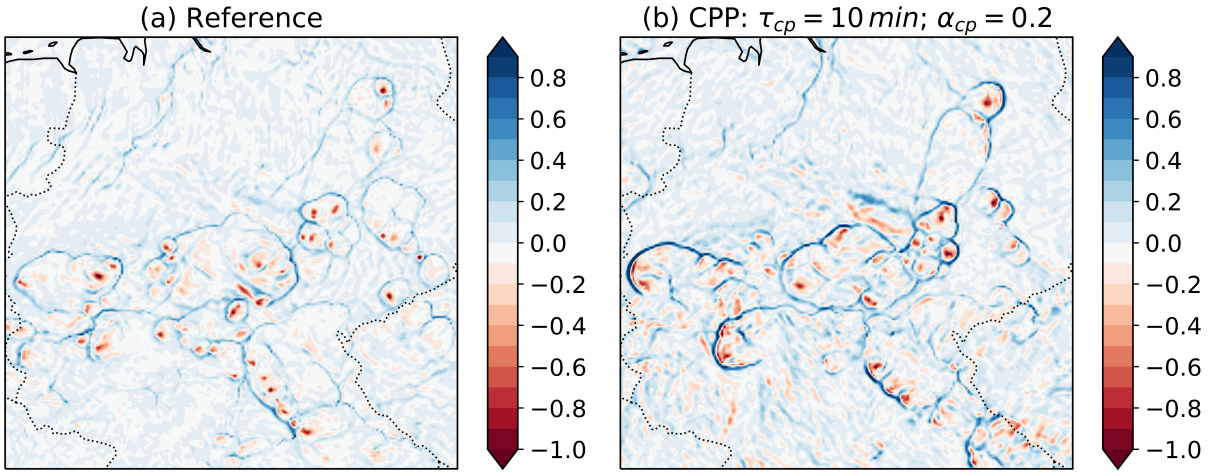


**Figure 41:** Vertical velocity for the reference (a) and a CPP (b) simulation is displayed. Fields are displayed for the 5th model level ($\approx 200\,m$) above surface at 15 UTC and a selected region with several cold pool gust fronts over central Germany.

### 6.3.2 *Impact on precipitation*

As the main incentive of CPP is to improve precipitation forecasts, we now evaluate its impact on precipitation in more detail. Considering example precipitation fields, we see the biggest differences in the late afternoon and evening. Fields for 19:00 UTC are displayed in Fig. 42 for illustration. Overall, there is more precipitation in CPP compared to the reference simulations, and the precipitation cells seem to be stronger, larger or more clustered together. Precipitation structures are also slightly more intense in CPP. [15] For more quantitative results, we will investigate both the impact on the diurnal cycle, i.e., afternoon/evening precipitation and the organization in a more quantitative way in the next two sections.

#### 6.3.2.1 *Diurnal cycle*

For the diurnal cycle, we first consider domain averaged precipitation amounts as a function of time, as displayed in Fig. 43a. Regarding the discrepancies between the model and observations in the first half of the day, the large scale situation (with regard to precipitation) is still captured by the model, with organized precipitation over southern Germany and the Alps. However, spurious convection early on, and too weak precipitation, later on, can explain the observed differences. The CPP schemes do not show any impact then. For our default setting of $\tau_{cp} = 10\,min$ we find an increased amount of precipitation starting around 13:00 UTC, which continues throughout the afternoon and evening. The most considerable differences occur between 15-18:00 UTC. CPP does not simply increase the amplitude of the precipitation peak, but a shift in precipitation towards the evening occurs. In comparison to the radar observations, these differences improve the domain averaged precipitation. While the precipitation peak

15 The stronger cells can be quantified via histograms of precipitation (not shown) and confirm that CPP has a higher frequency of extreme precipitation than both reference and radar. The differences to the reference in such histograms were, however, weak in relation to differences to the radar observations.

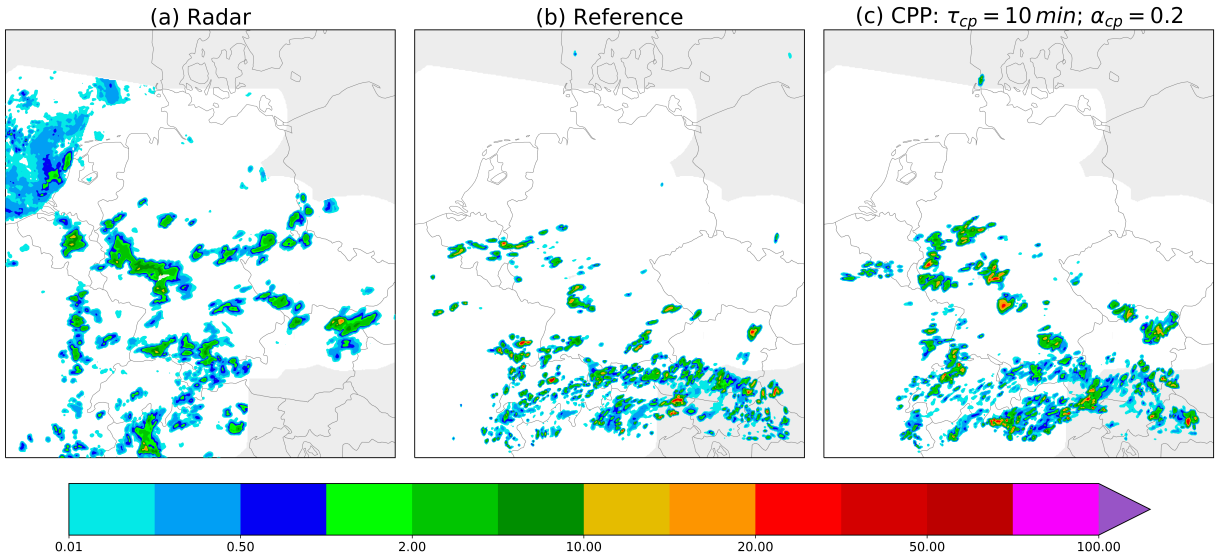


**Figure 42:** Example precipitation fields [$mmh^{-1}$] are shown for radar observations (a), the reference (b) and a CPP (c) simulation at 19 UTC.

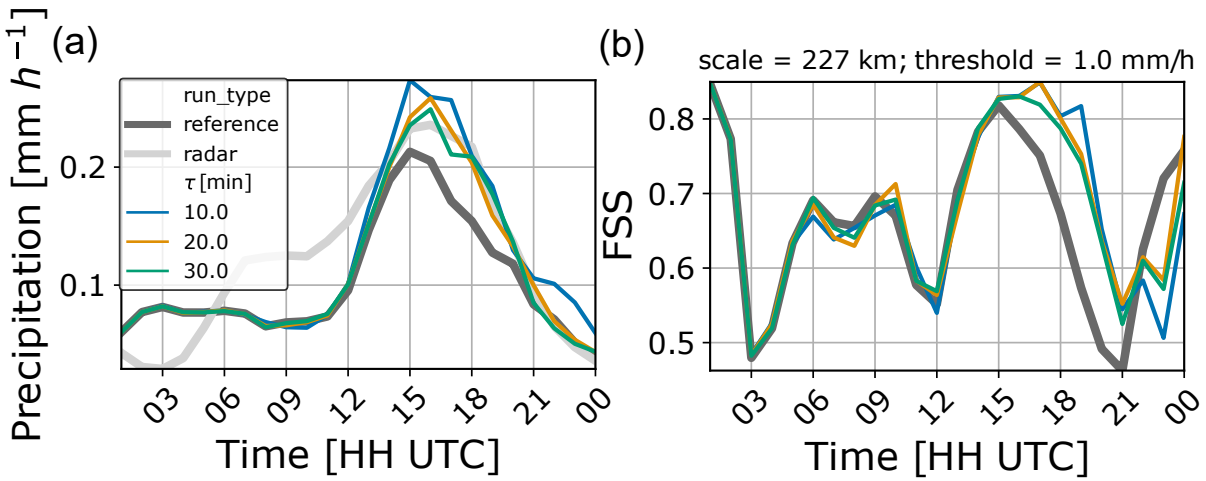


**Figure 43:** Time series of domain averaged precipitation (a) and fracion skill score (b). The bold dark grey lines display the reference simulation, the bold light grey line the radar observations and colored lines CPP configurations with different $\tau_{cp}$ but otherwise default settings. We have tested different scales and thresholds for the FSS as well, but the qualitative impact of CPP is not strongly sensitive for most settings.

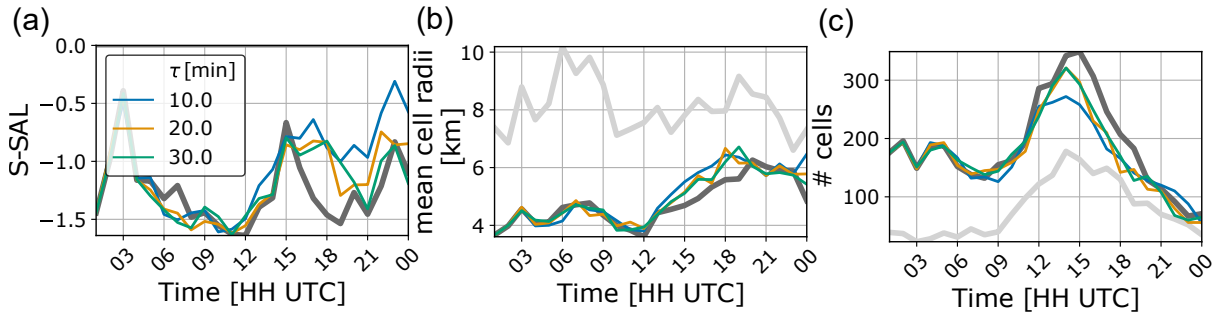


**Figure 44:** Time series of the structure component of the SAL score (a), average cloud radii (b) and cloud numbers (c). The bold dark grey lines display the reference simulation, the bold light grey lines the radar observations and colored lines the CPP configurations with different $\tau_{cp}$ but otherwise default settings.

around 15 UTC is somewhat overestimated by CPP (the reference underestimates the peak), the domain mean precipitation of CPP matches well with the observations from 18-21:00 UTC.

To verify that the enhanced precipitation by CPP is located somewhat realistically, we consider the Fraction Skill Score (FSS) displayed in Fig. 43b.[16] Considerable improvements to the reference simulation occur between 15-21 UTC, in accordance with the domain averaged precipitation amount. Improvements of up to 0.2 occur, e.g., at 19 UTC. After 22 UTC, however, the FSS of CPP deteriorates in comparison to the reference. We note that the displayed FSS considers a spatial scale of 227 $km$, which is relatively large for the convective scale. Smaller scales have been considered as well, which give similar responses of the CPP scheme, but with overall lower FSS values for all simulations.

We summarize that - in accordance with our objective - CPP enhances cold pool driven convective initiation and thereby specifically improves the afternoon and evening precipitation amount and forecast skill as quantified by the FSS.

#### 6.3.2.2 *Organization*

To quantify the organization of convection, we first consider the S(tructure) component of the SAL score in Fig. 44a[17]. Especially after 15 UTC less negative S-SAL values occur with the CPP scheme compared to the reference simulation. This confirms that cloud structures become larger and less peaked and thereby match the radar observations better. The mean cell radii computed from closed precipitating areas (Fig. 44b ) further confirm this aspect. They show that precipitation cells tend to be larger in the CPP simulation after 13 UTC. The values are, however, still too small compared to the observations. The total number of clouds also reduces after 12 UTC. We have further tested a watershed segmentation for treating merged cells separately. Differences to

16 The FSS compares the fractions of precipitating gridpoints within a specified neighborhood (scale) between model and radar data. Higher FSS values indicate more skill up to $FSS = 1$ as a perfect match. Values below $\approx 0.5$ are assumed to be without any skill. See Sec. 2.4.1 for more details.

17 The S(tructure) component of the SAL score is a quantitative measure for the structure of the precipitation field. Negative S values imply that the simulated precipitation cells are too small and peaked compared to the radar observations while $S = 0$ suggests a perfect match in terms of structure (see 2.4.2 for details).

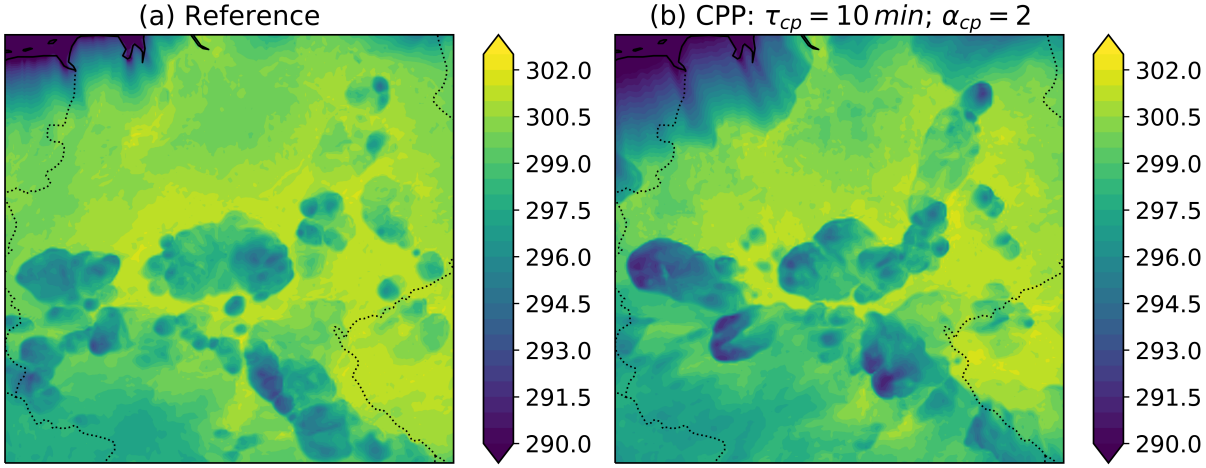


**Figure 45:** Virtual potential temperature $\theta_v$ is displayed for the reference (a) and a CPP (b) simulation for the lowest model level ($\approx 10\,m$) at 15 UTC. Note that only a subdomain of the whole simulation domain is shown.

the reference simulations were similar to the non-separated cells. To further quantify organization concerning cell distances, we considered the $I_{org}$ and radial distribution functions (Rasp *et al.*, 2018b; Pscheidt *et al.*, 2019). While the former evaluates the nearest neighbor distances, the latter considers cell densities at a given distance to one cell. Due to different cell size distributions, different total cell numbers, background gradients or simply noisy data, we struggled in interpreting the results and decided not to include them here.

In summary, there seems to be some form of stronger, area-based organization with CPP resulting in improved cell structures (S-SAL), larger single cell sizes and also larger, but fewer cell agglomerations. However, we could not identify a clear signal in the distance-based metrics for organization.

### 6.3.3 *Other impacts of CPP*

We have further found an impact of CPP on aspects other than precipitation, namely cold pools, land-sea breezes and the horizontal wind fields.

As we have speculated in the introduction, enhanced convective initiation at cold pool gust front may also be able to strengthen the cold pools themselves. We shortly evaluate this possibility by displaying example fields of virtual potential temperature $\theta_v$ at the lowest model level ($\approx 10\,m$) in Fig. 45. We can indeed confirm that many cold pools tend to be somewhat colder in the CPP simulations.

Fig. 45 further shows a stronger land-sea breeze near the coast, likely related to the application of the cold pool perturbations to gust fronts of the land-sea breeze. Land-sea breezes can also be regarded as density currents and can initiate new convection. Hence, the application of CPP to land-sea breezes can be reasonable as will be discussed in more detail in Sec. 6.4.

We have further identified a considerable impact on the horizontal wind fields. Example fields reveal substantial differences in the large scale structure of the wind field

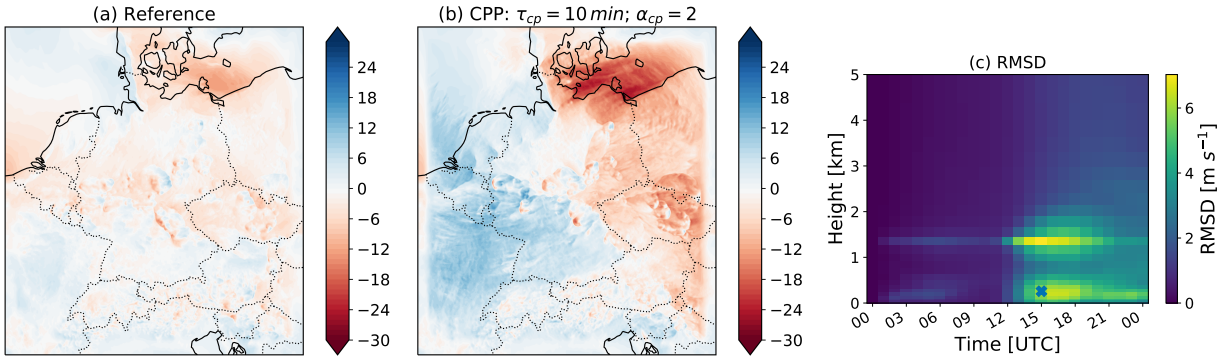


**Figure 46:** U-component of the horizontal wind speeds for the reference (a) and a CPP (b) simulation is displayed for the 6th model level ($\approx 300\,m$) above surface at 15 UTC. The root mean square difference (RMSD) of the horizontal, zonal wind (u) between the two simulations is displayed in (c). The cross marks the height and time for the fields in (a) and (b). As the RMSD above two kilometers is small, only the lowest five kilometers are shown.

(see Fig. 46a, b). By computing root mean square differences between the reference and the CPP simulations, as shown in Fig. 46, this impact is mostly evident in the lowest two kilometers in the afternoon. The two observed maxima in the vertical correlate well with the heights of the strongest horizontal wind perturbations, where the vertical gradient of the vertical velocity is strongest (not shown). Similar observations can be made for the meridional wind component $v$. We note, however, that above 2 km, the large scale wind field in the CPP run is reasonably similar to the reference simulation. Further discussion on possible causes, impacts and solutions of this horizontal wind impact will be given in Section 6.4.

### 6.3.4 *Parameter sensitivity of CPP*

As displayed in Table 3, several parameters have to be chosen within CPP. We shortly evaluate the impact of changes in four selected parameters, namely $\alpha_{cp}$, $\tau_{cp}$, $w^*_{max}$ and $k_{ref}$. Changes in the other parameters are expected to have only a small impact or an impact similar to one of the first four parameters: small changes in $\theta^*_{v,g}$, $n_{filter}$ or $\mu^*_{sso}$ will only marginally affect the results, as $w^*_{max}$ is the dominant constraint for masking the perturbations in the horizontal (see Fig. 40); changes in $\Delta H$ will likely behave similar to changes in $k_{ref}$, as in both cases, the perturbations will extent higher into the atmosphere.

SENSITIVITY TO $\tau_{cp}$ AND $\alpha_{cp}$ As the different simulations in Fig. 43 and 44 show, a reduction in $\tau_{cp}$ tends to increase the impact on precipitation, FSS, S-SAL, cloud sizes and numbers. This difference is more clear for $\tau_{cp} = 10\,min$ and $\tau_{cp} = 20\,min$, while the measures for $\tau_{cp} = 20\,min$ and $\tau_{cp} = 30\,min$ are more similar. For $\alpha_{cp} = 0.15$ (instead of 0.2), the response to changes in $\tau_{cp}$ are comparable (Fig. 47, dashed lines).

With regard to $\alpha_{cp}$, an increase in $\alpha_{cp}$ - resulting in larger target $w_0$ - seems to increase precipitation amount (e.g., for $\tau_{cp} = 30\,min$, green lines). If $\tau_{cp}$ is reduced too much, or $\alpha_{cp}$ increased too much, a stronger, nonlinear response occurs: for $\alpha_{cp} = 0.3, \tau_{cp} =$

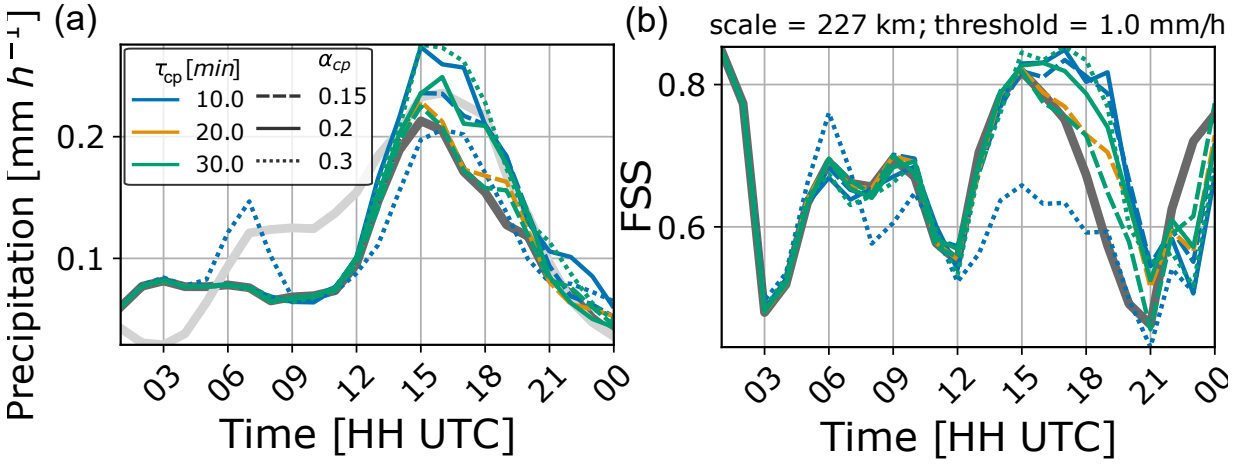


**Figure 47:** Time series of domain averaged precipitation (a) and fracion skill score (b) for different parameters of $\alpha_{cp}$ and $\tau_{cp}$. The bold dark grey lines display the reference simulation, the bold light grey line the radar observations and colored lines CPP configurations with different $\tau_{cp}$ and $\alpha_{cp}$ but otherwise default settings.

$10\,min$, i.e., the blue, dotted line, precipitation impact is already strong at 7:00 UTC but then reduced during the day joint with a strong loss of FSS. This suggests that the perturbations are too strong, resulting in spurious gravity waves, convection and other nonlinear responses that cause a loss of precipitation and skill. We have also considered S-SAL, cell sizes and numbers, but the behavior is mostly similar to the one displayed in Fig. 44 and hence not shown for different $\alpha_{cp}$.

Overall, these responses to changes in $\tau_{cp}$ correspond to our expectations: Reduced $\tau_{cp}$ will cause a faster imposition of the target $\Delta w$, hence larger perturbations per time step, and subsequently larger precipitation impact. Increases in $\alpha_{cp}$ result in larger target $w_0$ and hence also in larger perturbations per time step and a larger precipitation impact. If the perturbations are imposed too fast, or the target $w_0$ is too large, however, the simulations break down, nonlinear effects occur, and comparability to the reference simulation is lost.

SENSITIVITY TO $k_{ref}$ The precipitation response to changes in $k_{ref}$ - the height level up to which $w_{max}$ is searched for and the perturbations are imposed - is displayed in Fig. 48a. For an increase in height from level number 38 to 36 (from ≈1 km to ≈1.5 km), the impact increases. As the perturbations extend higher up, the probability of initiating new convection is likely enhanced. A further increase in height to $k_{ref} = 34$ (≈1.9 km), however, does not seem to enhance the precipitation response any further. We suggest, however, that $k_{ref}$ should be chosen as the expected height of the cold pool gust front.

SENSITIVITY TO $w^*_{max}$ The precipitation response to changes in $w^*_{max}$ - the minimum $w_{max}$ that is required for CPP - is displayed in Fig. 48b. If we increase $w^*_{max}$ to $0.7\,m\,s^{-1}$, fewer grid points will be perturbed, and we find a slightly reduced response in precipitation. If we reduce $w^*_{max}$ to $0.3\,m\,s^{-1}$, more grid points with weaker vertical velocities will be perturbed. Interestingly, for this situation, the overall precipitation amplitude is also reduced. However, deviations from the reference simulation occur already in the

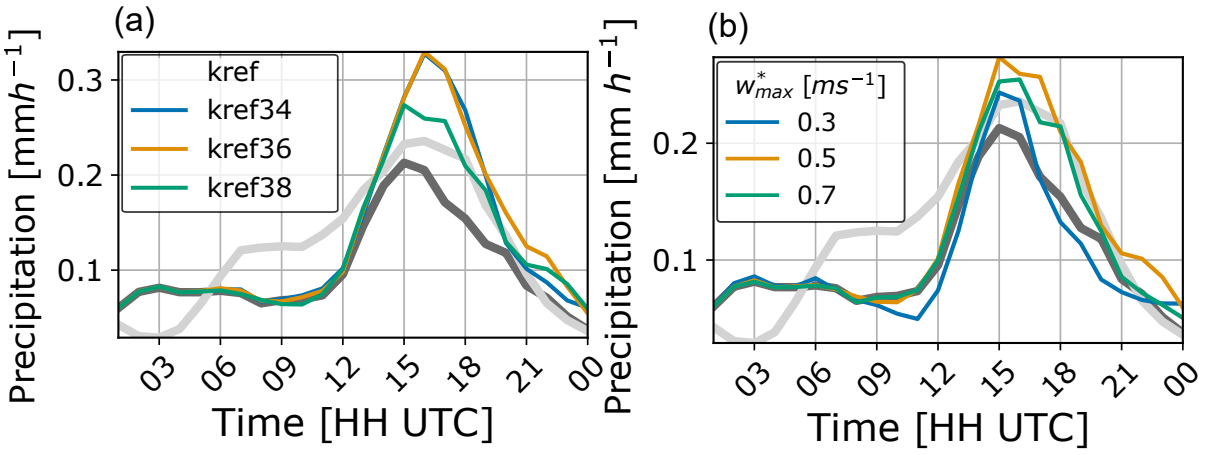


**Figure 48:** Time series of domain averaged precipitation for different simulations where (a) $k_{ref}$ and (b) $w^*_{max}$ is changed. The specified model levels 38, 36 and 34 for $k_{ref}$ correspond approximately to 1, 1.5 and 1.9 km heights. The bold dark grey lines display the reference simulation, the bold light grey line the radar observations and colored lines CPP configurations with different $k_{ref}$ and $w_{max}*$ but otherwise default settings.

morning, and a substantial reduction in FSS occurs (not shown). This is again an indication that the perturbations are too strong and acting at grid points, which may not be related to cold pool gust fronts. We conclude that $w^*_{max} \approx 0.5$ is a reasonable value, with a small sensitivity for higher values, but a break down of the CPP scheme with lower values.

### 6.3.5 *10 day period*

We now apply the CPP scheme on the ten days also used in Chapter 3 to evaluate the robustness of our results to other days and the adaptation of the scheme on different synoptic conditions. As in Chapter 3, we separate the ten days into five days with stronger synoptic forcing and five days with weaker synoptic forcing. The diurnal cycle of precipitation for the default CPP displayed in Fig. 49a (blue line) shows a similar impact as observed for 5 June 2016 for the weakly forced days: the precipitation is enhanced especially in from 15 to 19 UTC and thereby reduces the bias with regard to observations. Some small overestimation of precipitation now even occurs around 17 and 22 UTC. These are, however, small relative to the bias in the reference simulation. For the strongly forced period, the impact of CPP is small.

The FSS behaves similarly to the domain averaged precipitation (Fig. 49b). For weakly forced days, CPP improves the FSS from 14 to 23 UTC by up to 0.05. For the strongly forced days, the impact is small, but small improvements can be seen from 15 UTC onward. We have further evaluated the S-SAL score, cloud sizes and numbers, and the impact of CPP is similar to the impact on precipitation amount and FSS concerning the timing and the degree of improvements. An alternative CPP simulation with $\alpha_{cp} = 0.3$ and $\tau_{cp} = 30\,min$ behaves very similarly (light organge lines in Fig. 49).

We summarize that the impact of CPP on the weakly forced days is comparable to our findings for 5 June 2016, although somewhat weaker, while the impact on the

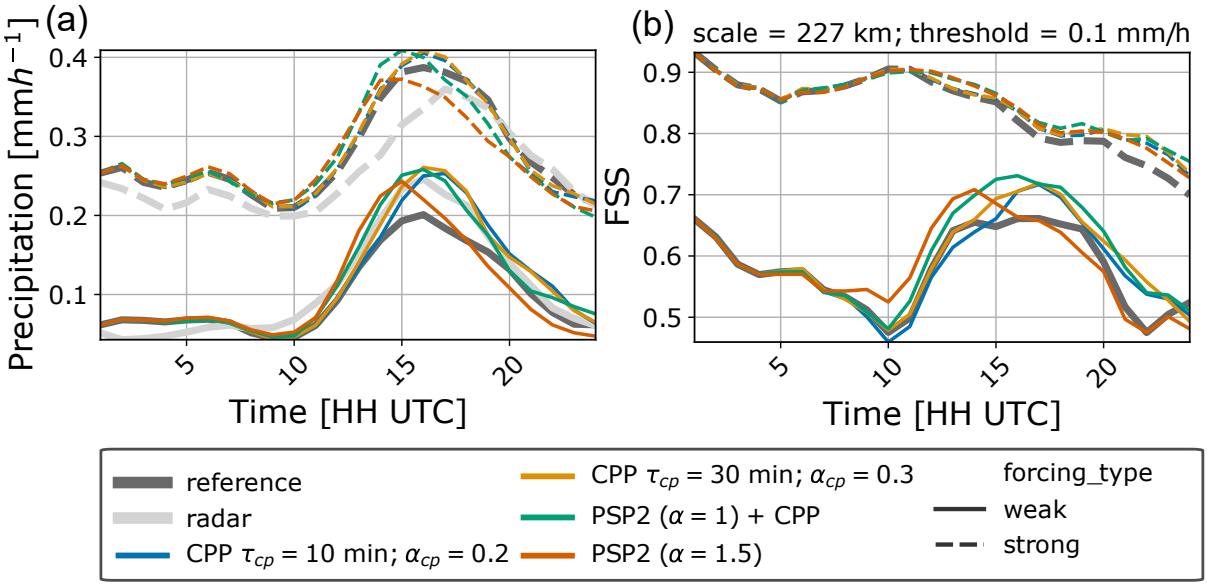


**Figure 49:** Diurnal cycle of domain averaged precipitation (a) and FSS (b) averaged over the four weakly and four strongly forced days. For the FSS, a precipitation threshold of 0.1 $mm\,h^{-1}$ and a scale of 227 km have been used. We have tested different settings as well, but the qualitative impact of CPP is not sensitive for most settings.

strongly forced days is small. Due to weaker surface heating or stratiform precipitation during the strongly forced period, we expect cold pools to be weaker and less frequent. Joint with weaker CAPE and stronger synoptic forcing for convection, cold pools are expected to be less critical for the initiation of convection during the strongly forced period. This can explain the weak impact of CPP for these days.

### 6.3.6 *Combining CPP with PSP2*

We now further evaluate the combined impact of PSP2 and CPP, denoted as PSP2+CPP. For comparison, we have also included the PSP2 run with the settings concluded from Chapter 3 in Fig. 49. For the combined simulation, however, we rescaled the amplitude of PSP2 from $\alpha_{tuning} = 1.5$ to $\alpha_{tuning} = 1$ to yield comparable precipitation amplitudes. From Fig. 49a we can indeed confirm, that both characteristics of PSP2 and CPP are visible in the combined simulation: the onset of precipitation is shifted slightly towards the morning in comparison to the CPP or the reference simulation; the afternoon/evening precipitation amount is increased towards more realistic values in comparison to the PSP2 or the reference simulation for the weakly forced period. For the strongly forced period, a considerable increase in precipitation amplitude is visible for the combined, PSP2+CPP simulation in comparison to the PSP2 simulation, although only a small increase is visible for CPP in comparison to the reference simulation. Considering Fig. 49b, the FSS of PSP2+CPP is considerably improved for the weakly forced period compared to the reference simulation and exceeds both CPP and PSP2 for most of the time.

## 6.4 DISCUSSION

In this chapter, we developed deterministic perturbations for the three wind components with the incentive to strengthen cold pool gust fronts and thereby enhance the afternoon/evening precipitation and organization of convection. The perturbations were designed so that the vertical velocity in the lowest kilometer at cold pool gust fronts converges towards a target vertical velocity, which was obtained from dimensional analysis and estimated by local buoyancy gradients.

### 6.4.1 *Benefits for precipitation*

Example fields of vertical velocity revealed stronger updrafts at cold pool gust fronts with CPP compared to the reference simulation. Afternoon/evening precipitation is enhanced, and FSS improved by CPP. Both S-SAL and cloud size and numbers improved to a small amount for the CPP simulations, which provide proxies for area-based organization. These results confirmed that CPP indeed achieves the original goals of the scheme.

These generally positive results were further confirmed on five weakly forced days, although to a lesser extent. The impact of CPP on five strongly forced days is weak, as expected, due to the assumed weaker role of cold pools for convective initiation. Hence, the design and settings of CPP, which we developed/chose based on 5 June 2016, are also applicable to other situations, and CPP adapts well to different situations.

Furthermore, we tested how sensitive the impact of CPP is to changes in some of the CPP settings. As the most direct tuning parameters of CPP, reducing the time scale $\tau_{cp}$ or enhancing $\alpha_{cp}$ results in stronger perturbations and impacts of CPP. We found further sensitivities to other parameters, but we recommend their default values based on physical reasoning. If the chosen settings result in unusually high perturbation amplitudes, i.e., by using too small $\tau_{cp}$, $w^*_{max}$ or too large $\alpha_{cp}$, spurious gravity waves, convection and other artifacts arise, and the simulations lose their comparability to the reference simulation. Such settings should be avoided. We hope this sensitivity study will give some guidance for future applications with CPP on which settings to choose.

The combined impact of CPP with the PSP2 scheme provides an additional highlight, as the combined simulation outperformed both singularly perturbed simulations. The results confirm the complementary behavior of the two process-based perturbation schemes PSP2 and CPP, and hence also the relevance of boundary-layer turbulence and cold pools for initiating convection. We can envision modeling frameworks where a multitude of complementary, process-level perturbation schemes can be combined to improve weather forecasts. Such frameworks may also enable a better evaluation of the relevance of single processes or their interactions.

### 6.4.2 *Other impacts of CPP*

Aside from precipitation, we also identified impacts of CPP on other aspects, namely on cold pools, land-sea breezes and horizontal wind fields.

Example $\theta_v$ fields near the surface suggest that cold pools are stronger with CPP. A proper quantification of this effect is currently addressed in a Bachelor thesis, where the cold pool detection algorithm from the previous chapter (Chapter 5) is applied to the here used COSMO simulations. Preliminary results indeed confirm this indication. Such strengthening of the cold pools by CPP hints at a feedback loop by which stronger cold pool driven convective initiation by CPP strengthens the convection within the original cold pools and thereby the cold pool itself. This feedback loop has further been mentioned by Böing *et al.* (2012); Schlemmer and Hohenegger (2014).

These $\theta_v$ fields further displayed an impact on the land-sea breeze, where CPP was also active. As the land-sea breeze is also a form of a density current (e.g., Moncrieff and Liu, 1999), we expect a similar resolution sensitivity as for cold pools. In addition, land-sea breezes can also initiate new convection. These similarities between cold pool gust fronts and land-sea breezes suggest that it may be indeed appropriate to apply CPP to land-sea breezes. Further investigation will be necessary to identify whether land-sea breezes are indeed insufficiently resolved at km-scales and whether the CPP scheme is an appropriate solution. This includes the applicability of the CPP settings, such as the height scale $H = 200\,m$ for land-sea breezes.

We further found an impact of CPP on the horizontal wind fields in the lowest two kilometers where the w-perturbations change strongest with height. The structure of this impact suggest that the positive, vertical velocity perturbations require a convergent flow near the surface, where the w-perturbations increase with height, and a divergent flow at $\approx 1.5\,km$, where the w-perturbations decrease with height to satisfy 3d non-divergence. Due to the very localized w-perturbations, we expected the compensating horizontal flow to be small. Unfortunately, this seems not to be the case. Nonetheless, we do not expect that this behavior has a strong effect on precipitation: the strength of the vertical velocity should be the primary mechanism for convective initiation, and, as the effect is localized to few height levels in the boundary layer, convective updrafts in the free troposphere should not be affected. Also, the positive impact on the spatial distribution of precipitation implies that the large scale situation is still sufficiently represented. For the operational application of CPP, however, other aspects, e.g., transport of pollutants in the boundary layer, have to be considered as well, which may be negatively affected by the current impact of CPP on the horizontal wind fields. Hence, a reduction of this impact is desireable for the future. A possible solution could be to include compensating subsidence nearby cold pool gust fronts.

### 6.4.3 *Future steps*

Above, we already described possible actions to improve the current CPP scheme, most importantly, addressing the impact on the horizontal wind. Two other steps that should be carefully considered are the modification of CPP for scale adaptivity and a comprehensive tuning of the model joint with CPP.

As scale adaptivity is a crucial feature for grey zone parameterizations (see Sec. 1.2.1), it is worth considering the applicability of the current CPP for different model resolutions. Certainly, CPP should only be applied if the models can explicitly resolve cold pools and deep convection. This likely limits CPP to resolutions below ten or even

five kilometers. Nevertheless, the parameters chosen here will likely be required to be adapted. $w^*_{max}$ is expected to depend on model resolution as vertical velocity generally depends on it (Weisman *et al.*, 1997; Pauluis and Garner, 2006; Morrison, 2016a,b; Jeevanjee and Romps, 2016; Jeevanjee, 2017). Theoretically derived relationships for $w$ and grid size, however, already exist (e.g., Jeevanjee, 2017) and could be used to formulate a scale-dependent $w^*_{max}$. Similarly, the threshold to exclude orographic grid points $\mu^*_{sso}$ could be formulated using scaling laws of orography. As cold pool strengths also appear to be sensitive to resolution (see chap. 5), gradients of $\theta_v$ will change as well. Despite the expected, weak sensitivity of the CPP impact to $\theta^*_{v,g}$, the chosen threshold for $\theta^*_{v,g}$ may have to be adapted for different model resolutions. If no theoretical formulation can be found, empirical relationships could be used to find a scale adaptive formulation of $\theta^*_{v,g}$. Simulations with different model resolutions will then allow insight on the scale adaptivity of CPP.

We emphasize that current models have been tuned over years - and sometimes even decades - to best fit observations. The arising presence of compensating errors is widely acknowledged (e.g., Palmer and Weisheimer, 2011; Mauritsen *et al.*, 2012; Hanley *et al.*, 2015; Berner *et al.*, 2017; Hourdin *et al.*, 2017), and challenges the development of process-based schemes such as CPP. One aforementioned example includes the tuning of the turbulent lenght scale `tur_len` to improve precipitation forecasts at the expense of temperature biases (Baldauf *et al.*, 2011; Schraff *et al.*, 2016). By providing process-based schemes instead, we hope that the physical processes will be represented more realistically, and resulting model responses will occur for the right reasons. Another approach that could prove beneficial with regard to cold pools is the use of a two-moment microphysics scheme, which allows for a more sophisticated representation of cloud condensation and evaporation processes. But to obtain a useful model setup with CPP and a two-moment microphysics scheme, parameters of both schemes have to be tuned simultaneously. Ideally, such process-based schemes, including also PSP2, should be incorporated in the comprehensive tuning process of operational models before they can be used operationally. This is a costly endeavor, beyond the scope of any researcher. We hope we have given enough reasons for national weather centers to take over.

In summary, to address our posed research goals, we successfully developed a cold pool parameterization which strengthens cold pool gust fronts and - to all appearances - thereby enhances cold pool driven convective initiation. As anticipated, the late afternoon/evening precipitation improves for weakly forced situations, and enhanced organization is indicated. Further development and evaluation may be necessary to ensure scale adaptivity and to quantify unforeseen side effects.

# 7

# SUMMARY AND CONCLUSIONS

## 7.1 SUMMARY

Some of the largest problems of convective precipitation forecasts in state-of-the-art, convection-permitting models are biases in the diurnal cycle and the organization of convection. A large portion of theses deficits is currently ascribed to the insufficient representation of processes leading to convective initiation. These processes occur in the boundary layer and are mostly too small to be sufficiently resolved even by convection-permitting models. This thesis has contributed to this unsettled research topic - regarding the deficiency in convective initiation - by identifying model deficits and potential solutions. Specifically, three of the most important processes for convective initiation have been considered in detail, namely boundary-layer turbulence, subgrid-scale orography and cold pools.

REPRESENTING BOUNDARYLAYER TURBULENCE VARIABILITY Turbulence is a ubiquitous process in convective boundary layers and tightly coupled to convective initiation. Km-scale models traditionally include turbulence in deterministic, parameterized form, i.e., the mean impact on the resolved scales is considered. As the subgrid variability of turbulence can be crucial for convective initiation, the PSP scheme was introduced by Kober and Craig (2016) to account for this subgrid variability in a stochastic manner and to enhance convective initiation. In the first part of this thesis, four modifications of this scheme were considered to improve its physical consistency and reduce unwanted side effects. These included an autoregressive, continuously evolving random field; a limitation of the perturbations to the boundary layer that removes artificial convection at night; a mask that turns off perturbations in precipitating columns to retain coherent structures; and non-divergent wind perturbations that drastically increase the effectiveness of the vertical velocity perturbations. In a revised version, PSP2, the combined modifications retain the physically-based coupling to the boundary-layer scheme while being physically more consistent. In addition, the spatial structure of the simulated precipitation cells was improved and spurious night-time precipitation reduced. Overall, this scheme has the potential to improve predictions of convective initiation in kilometer-scale models while minimizing other biases.

REPRESENTING LIFTING BY SUBGRID-SCALE OROGRAPHY Orography is an undeniable trigger for convective initiation by a range of different mechanisms. In km-scale models, a large part of the orography is explicitly included, and the relevance of small scale, i.e., subgrid-scale orography, for convective initiation is ambiguous. Nonetheless, we developed a stochastic scheme to include the effect of mechanical lifting by subgrid-scale orography, the SSOSP scheme. The mechanical lifting effect was chosen as the

most direct effect and approximated by gravity wave formalism and information on subgrid-scale orography. While a clear increase in convective initiation over orographic regions was possible by the SSOSP scheme, this usually coincided with an undesired impact also on non-orographic regions. We concluded, that too large perturbation amplitudes were necessary to see a substantial impact and that most likely, subgrid-scale orography is not of primary relevance in the presence of resolved orography. Hence, we will focus from now on more on PSP2 and cold pools.

In both the PSP2 and the SSOSP stochastic perturbations were chosen to better represent subgrid-scale distributions in the grey zone. This combination of process-level perturbations and stochastic components is rare and innovative. It enables the reduction of model errors by noise-induced drift, for instance by improving precipitation amplitudes or the diurnal cycle of precipitation; but also it facilitates the quantification of model uncertainty at its source, which is crucial - yet uncommon - for reliable ensembles. For example, the commonly used stochastic approaches SPPT (Buizza *et al.*, 1999) and SKEBS (Shutts, 2005) only facilitate the increase in ensemble spread without any attribution to specific, physical processes.

COLD POOL DRIVEN CONVECTIVE INITIATION The two most notable chapters of this thesis are concerned with cold pool driven convective initiation. Cold pools provide triggering mechanism nearby already existing convection and thereby contribute to the organization of convection. This further can prolong the lifetime of convection and be an effective trigger mechanism in the late afternoon, which influences the diurnal cycle of convection. Whether cold pools and their contribution to convective initiation are properly represented in km-scale models was an open question. Hence, cold pool driven convective initiation was investigated in the first of the two cold pool chapters with a focus on the diurnal cycle and organization of convection and the sensitivity to grid size. Simulations of the ICON-LEM model with grid sizes from 156 m to 625 m over Germany were used. In these simulations, we identified cold pools, cold pool boundaries and initiated convection. Cold pool driven convective initiation can be as dominant as 50% of total convective initiation, in particular in the late afternoon. In addition, cold pool related triggering is many times more efficient than without cold pools. By comparing different model resolutions, we found that cold pools are more frequent, smaller and less intense in lower resolution simulations. Furthermore, their gust fronts are weaker and less likely to trigger new convection. To identify how model resolution affects this triggering probability, we used a linear causal graph analysis. In doing so, we postulated a graph structure with potential causal pathways and then applied multi-linear regression accordingly. Despite several possible indirect pathways, we found one dominant, systematic effect: reducing grid sizes directly reduces upward mass flux at the gust front, which causes weaker triggering probabilities. These findings are expected to be even more relevant for km-scale, numerical weather prediction models. We expected that a better representation of cold pool driven convective initiation will advance forecasts of convective precipitation.

Consequently, in the second cold pool chapter, we developed a deterministic cold pool scheme to strengthen cold pool gust fronts and thereby improve cold pool driven convective initiation in km-scale models. A target vertical velocity scale for cold pool

gust fronts was derived using dimensional analysis and approximated based on local buoyancy gradients. Vertical velocity perturbations were then imposed so that the model vertical velocity at cold pool gust fronts converges towards this target scale. In addition, horizontal wind components were perturbed in a 3d non-divergent manner as for PSP2 and SSOSP. Our results showed that cold pool gust fronts are strengthened by CPP and - to all appearances - enhances cold pool driven convective initiation. As anticipated, the late afternoon/evening precipitation is improved for weakly forced situations and indications for enhanced organization were found. Further development and evaluation may be necessary to ensure scale adaptivity and to quantify and eliminate side effects.

In addition, we evaluated the combined use of both PSP2 and CPP, as they represent two complementary processes. Such a combined simulation revealed the benefits of both schemes, with an earlier onset of convection and improved afternoon/evening precipitation. This confirms the possibility, that process-based schemes can be used in an additive way to yield quasi-additive improvements of the forecasts.

## 7.2 FUTURE STEPS FOR PSP2 AND CPP

For now, we have focused on understanding the systematic changes caused by the perturbation schemes rather than on absolute forecast scores. Thus, whether PSP2 or CPP[18] will improve forecasts still needs to be fully determined before they can be used in operational weather forecasts. To do so, the following steps to further evaluate and develop the schemes are advisable.

1. As the operational use and development of the here used COSMO-DE model will be migrated to the ICON model in the future (personal communication with DWD), so should PSP2 and CPP. First pursuits to transfer PSP2 to the ICON model are currently underway, but new challenges due to an irregular grid structure in ICON arise.

2. The computational efficiency of the schemes will likely have to be improved. This can be achieved by improving the computational implementation of the schemes, or by allowing approximations of some of the more cost-intensive aspects of the schemes. This will be a priority for implementing PSP2 in ICON.

3. The schemes should be tested at different model resolutions to evaluate their scale adaptivity, a greatly desirable characteristic (see Section 1.2.1). Additional development will likely be required to satisfy it.

4. Simulations using a full ensemble prediction system over a sufficiently long time period are necessary to thoroughly validate the impact of the schemes. This should include data assimilation, a comprehensive selection of different weather situations, and comparison to observations.

5. A retuning of other model parameters will likely be necessary. Dealing with a highly complex model with dozens of adjustable parameters inevitably conjures

18 As we have not found any potential benefit of the SSOSP scheme, I will concentrate on PSP2 and CPP.

up the curse of model tuning (Palmer and Weisheimer, 2011; Mauritsen *et al.*, 2012; Hourdin *et al.*, 2017). Over many years the parameters in an operational model are modified to achieve the best forecast scores. This most likely invoked a multitude of compensating errors. Introducing more physical parameterizations, therefore, demands a careful retuning of the parameters.

Most of these steps are of technical nature and should be addressed by the capacities and expertise of national weather services. We hope that, by now, national weather services have been convinced of the potential benefits of PSP2 and CPP to continue further development. In fact, the promises of the PSP scheme have motivated the German, Swiss and British weather services to adapt similar schemes for testing (personal communications with DWD, MeteoSwiss and MetOffice). Also, in a current project by DWD, the SINPHONY project (Deutscher Wetterdienst, 2020), plans exist to integrate PSP2 in their new setup (personal communications).

## 7.3 CONCLUSIONS

The research presented in this dissertation contributes to open research topics in different ways. We provided answers to the open research questions concerning the subordinate relevance of small scale orography for convective initiation or the resolution sensitivity of cold pools and cold pool driven convective initiation. Such improved understanding helped us, but will likely also help others in directing their research. Furthermore, we developed three different schemes, PSP2, SSOSP and CPP, which provide novel approaches to address boundary-layer turbulence, subgrid-scale orography, cold pools, convective initiation and model uncertainty. The specific advantage of PSP2 lies in the combination of a stochastic component and its process-level, physically-based design, which not only reduces biases but also allows the quantification of uncertainty at its source. The CPP scheme provides another, notable innovation on how cold pools can be better represented at grey zone resolutions. By combining both PSP2 and CPP we confirmed their quasi-additive benefits, which helps to improve convective initiation in a more realistic, flow-dependent way, without the requirement to compensate errors with other processes and tuning. Importantly, PSP2 and CPP showed clear positive impacts on precipitation.

By developing these schemes, we substantially improved the representation of convective initiation in convection-permitting models. We comprehensively included processes relevant for convective initiation. At this point, we do not expect further, measurable benefits of including more processes or developing different schemes. Land-sea breezes, for instance, are already partially addressed by the CPP scheme, and other processes will likely only marginally impact convective initiation. Nonetheless, not all biases in convective precipitation forecasts were eliminated by the improved representation of convective initiation in this thesis. For further improvements in convective precipitation, aspects other than convective initiation should now be considered. The currently most promising aspects are better representations of microphysical processes, advancing convective-scale data assimilation and making good use of ensemble simulations.

Microphysical processes, such as evaporation, condensation and droplet formation, are often represented by single-moment microphysics schemes and are likely the most crudely approximated processes in current numerical models, despite their relevance for almost any weather situation. Using and developing two-moment microphysics schemes, instead, allows a more sophisticated representation of these processes, which may prove useful for the representation of clouds, convection and precipitation in the future (Seifert and Beheng, 2006; Baldauf *et al.*, 2011; Igel *et al.*, 2015).

Convective scale data assimilation uses a growing network of convective scale observations in combination with previous forecasts to obtain better initial conditions for current forecasts. Due to its youth (see, e.g., Schraff *et al.*, 2016), its potential possibilities are not yet fully exploited (Gustafsson *et al.*, 2018). In addition, data assimilation further requires reliable ensemble spread of the previous forecast, which is often missing and hence artificially inflated. Zeng *et al.* (2020), for instance, showed that the use of the PSP scheme within data assimilation cycling improves forecasts due to enhanced spread. Further development of data assimilation techniques and their combination with new techniques to quantify model uncertainty will likely benefit forecasting convection.

This already emphasizes the need for reliable ensembles. Most fundamentally, ensembles will allow the simulation and quantification of uncertainty. Then, the limits of predictability can be detected and the reliability of the forecast quantified. Unfortunately, finite computational capacities do not allow for sufficiently large ensembles for operational weather forecasts. This leads to under-dispersive ensembles and spurious correlations for data assimilation. Necker *et al.* (2020), however, has shown, that such sampling errors can be reduced for small ensembles. Further research in improving convective scale model spread, in correcting sampling errors and in determining necessary ensemble sizes will benefit forecasts of convective precipitation.

In this dissertation, we have comprehensively improved the representation of convective initiation in convection-permitting models. This contributed to improved representation of convection itself and, thereby, will potentially benefit forecasts of precipitation in the future. Such improved forecasts will allow more reliable and early warnings and can help our society to make better decisions to reduce casualties, financial losses and to improve our quality of life.

## 7.4 CAUSAL METHODS AND THEIR POTENTIAL FOR IMPROVING NWP

This thesis contributed to the research community also on a methodological level. We provided an example, how causal methods can be used with atmospheric modeling data to better understand the simulation of underlying processes (cold pool driven convective initiation). This gained understanding then helped to build a better parameterization (the CPP scheme). We believe, such a procedure has great potential to improve the development of parameterizations and NWP in the future.

The formal theory of causality is still fairly new to atmospheric sciences and only a small - but growing - number of applications exist. Recently, Nowack *et al.* (2020) used causal networks to evaluate climate models and Krich *et al.* (2020) to better understand biosphere-atmosphere interactions. The main advantage of the causal methods applied in this thesis lies in the formal framework, which specifies how causal effects can be

estimated from non-interventional data without the need for complex methods. Other approaches even enable the detection of causal structures, i.e. causal discovery (Pearl, 2009). Developments of new approaches and generalizations are growing. Runge *et al.* (2019), for instance, introduce an approach to detect nonlinear causal relationships in time series data.

These causal methods are useful for observations, but also simulation data sets where focused interventions are often not feasible. The increasing availability of computationally expensive model output will provide new opportunities to explore the physical processes controlling cumulus convection and many other weather phenomena across a wide range of scales. Such data sets include the high-resolution simulations made available by the HD(CP)$^2$ project (Heinze *et al.*, 2017; Stevens *et al.*, 2019a) and the DYAMOND model intercomparison project (Stevens *et al.*, 2019b), or very large ensembles as in Necker *et al.* (2020). The gained process understanding can then be used, for instance, to build better parameterizations.

Nonetheless, for high-dimensional, causal structures or interacting, nonlinear causal relationships, the application of such causal methods can be complex, if not impossible. Current developments of combining causal approaches with machine learning techniques seem promising and may overcome some fundamental problems of both causal and machine learning methods (see e.g. Athey, 2017; Schölkopf, 2019; Pearl, 2019). The successful outcome of such developments will have great implications not only for numerical weather prediction.

# APPENDIX

# A

# DETAILS ON THE CAUSAL ANALYSIS

## A.1 SINGLE DOOR CRITERION FOR THE CAUSAL MODEL

We provide the outcome of applying the single door criterion (see Sec. 5.1) to our causal structure (Fig. 35). Each set of adjustment variables that are necessary to estimate the direct causal effects by means of linear regression is given in Table 4. The D in brackets for the RI path marks a special situation: It is not necessary to include the diurnal/synoptic variables for the RI path, because we have postulated that model resolution does not impact the large scale situation. However, including it should not affect the regression estimate either, unless this hypothesis is violated. In the following section, we will also evaluate this hypothesis by comparing the regression estimates of RI when the diurnal/synoptic conditions (D) are included or excluded.

## A.2 SENSITIVITY AND ROBUSTNESS OF THE CAUSAL ANALYSIS

In this section we test the robustness of our causal analysis to some of the more important assumptions used in constructing the graph.

EVALUATION OF LINEAR REGRESSION By evaluating some standard properties of the applied linear regressions, some weak conclusions for our causal model can be drawn. We focus here on the p-values, which contain similar information as the bootstrapping, and the coefficients of determination $R^2$ (Wilks, 2005).

The p-values, displayed in Fig. 50a, b, are mostly small (below 0.05), which suggest that the estimated correlation coefficients differ significantly from zero. This implies that the corresponding variables, that we have identified, are indeed correlated. It follows that the variables are not d-separated under the adjustment sets given in Table 4. Consequently, there must be some causal connection between the selected variables. This causal connection, however, does not have to correspond to our causal structure (Pearl, 2009). Note that the RP and BP path (Fig. 50a) and a few path coefficients for single days (Fig. 50b), do not yield small p-values and no causal connection can be inferred.

The $R^2$ values are given in Fig. 50c, d and characterize the relative amount of variability that can be explained by the predictors. For gust front mass flux and triggering probability as predictands, the $R^2$ values are approximately 0.4 (Fig. 50c), which implies that a substantial part of their variability is included in our model. If separate days are considered, the highest $R^2$ values are ca. 0.8 (Fig.50d). Such high values confirm that a substantial part of the information necessary to determine the predictor has been included. It is consequently less likely that we have missed influential common drivers. Unfortunately, not all cases show such high $R^2$ values. This can be problematic,

**Table 4:** Estimation of direct path coefficients: To allow the estimation of path coefficients by multiple linear regression estimates, adjustment sets need to be included. The necessary adjustment sets are given by the single door criterion. They are then used as additional predictors (together with the parent of the direct path) within the multiple linear regression for predicting the child of the given path. For the RI path, the adjustment set [D] is not strictly necessary, since no link between RD is assumed.

| Path | | | adjustment set |
|---|---|---|---|
| R | -> | I | [ D ] |
| R | -> | B | I, D |
| I | -> | B | R, D |
| R | -> | G | I, B, D |
| I | -> | G | R, B, D |
| B | -> | G | R, I, D |
| R | -> | P | I, B, G, D |
| I | -> | P | R, B, G, D |
| B | -> | P | R, I, G, D |
| G | -> | P | R, I, B, D |

if the missing informative variables also influence other variables of our causal model (common driver).

STRATIFICATION BY MODEL RESOLUTION We can further evaluate the causal effects by using only two of three model resolutions. Corresponding causal effects are displayed in Fig. 51. We find that no substantial differences exist whether 156m - 312m, 312m - 625m or 156m - 625m are compared. Hence, within the given range of model resolution, the sensitivity to model resolution is indeed sufficiently linear.

STRATIFICATION BY DAYS As described in Sec. 5.2 and 5.3, the causal effects can also be evaluated for each day. The high variability in the RP path implied that the large scale conditions influence the RP effect, violating our linearity assumption. Since this is an important sensitivity it is discussed extensively in the main text in Chapter 5.

SENSITIVITY OF CHOSEN SYNOPTIC VARIABLES We have specified six synoptic variables that we use to describe the synoptic and diurnal situation. We have also considered several other variables (e.g., specific humidity or virtual potential temperature) and used each variable separately. These modifications in the synoptic variables however, did not lead to any substantial deviations in the estimated direct and indirect path coefficients from model resolution to triggering probability (not shown).

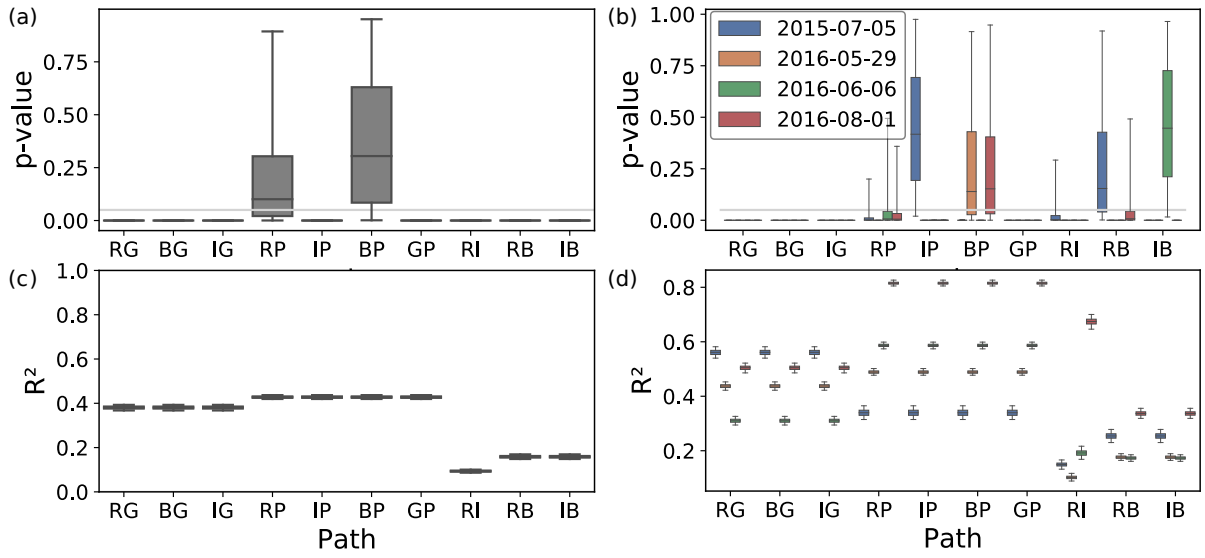


**Figure 50:** The p (a, b) and $R^2$ (c, d) values of the estimated path coefficients. (a, c) for all cold pools and (b, d) when stratified according to the four different days. The whiskers of the box plots correspond to the 95% confidence intervals from bootstrapping.

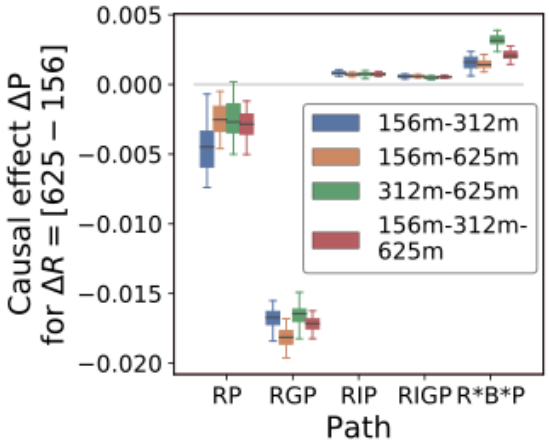


**Figure 51:** Aggregated path coefficients computed using only two of the three model resolutions. The whiskers of the box plots correspond to the 95% confidence intervals from bootstrapping.

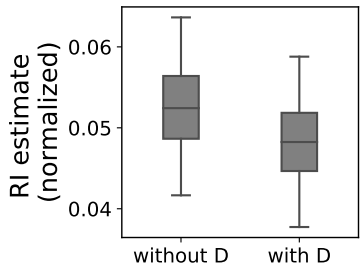


**Figure 52:** Normalized path estimates for the impact of model resolution on cold pool intensity (RI path) by including or excluding the diurnal/synoptic variables (D) as adjustment set for the linear regression. The whiskers of the box plots correspond to the 95% confidence intervals from bootstrapping.

DIFFERENT VARIABLES The choice of variables used to specify the I- and G-node is not unambiguous. Consequently we have evaluated several reasonable possibilities. We have determined the I-node by the area-averaged intensity, area-integrated intensity or the effective radius of the cold pools. To measure the gust front strength, we have considered 95th percentile vertical velocity and area-averaged or -integrated upward mass flux within the cold pool boundaries. The different combinations of variables however did not substantially impact the results of the causal path estimates (not shown).

IMPACT OF RESOLUTION ON SYNOPTIC CONDITIONS Properly evaluating the effect of model resolution on the diurnal/synoptic conditions (RD) is not trivial, because links between the different large scale variables are to be expected (this complicates the identification). An indication that this RD effect is small is provided by the small, mostly unsystematic differences between resolutions of each large scale variable (see Fig. 26). Additionally, we can include or exclude the large scale variables when estimating the RI path. If there is indeed no link between model resolution and diurnal/synoptic conditions, both estimates will be equal. The corresponding two estimates for the RI path are displayed in Fig. 52. The strongly overlapping distributions of the bootstrapping sample suggest, that the two estimates are not significantly different. Thus, it is reasonable to neglect a causal connection between model resolution and synoptic conditions.

# B

# DIMENSIONAL ANALYSIS FOR COLD POOL GUST FRONTS

Here, we give the full derivation for a characteristic vertical velocity scale $W$ for cold pool gust fronts, which is used in Chapter 5. We consider the inviscid, Boussinesque approximated set of equations for the streamfunction $\Phi$, vorticity $\eta$ and vertical velocity $w$ in a two dimensional plane in x and z, which is often considered for density currents (Rotunno *et al.*, 1988; Weisman and Rotunno, 2004; Bryan and Rotunno, 2014b):

$$\begin{aligned}
\frac{d\eta}{dt} &= -\frac{\partial b}{\partial x}, \\
\nabla^2\Phi &= \eta, \\
w &= -\frac{\partial \Phi}{\partial x} \text{ and} \\
w &= \frac{dz}{dt}.
\end{aligned}$$

We consider a parcel situated at a horizontal, time-independent buoyancy gradient $\frac{\partial b}{\partial x}$ (i.e., a cold pool gust front). First, we non-dimensionalize the equations using characteristic scales $B$, $H$, $L$, $W$, $\eta^*$, $\Phi^*$ and non-dimensional variables denoted wtih a tilde-sign, e.g. $\tilde{x}$:

$$\begin{aligned}
b &= B\tilde{b}, \\
x &= L\tilde{x}, \\
z &= H\tilde{x}, \\
t &= T\tilde{t}, \\
w &= W\tilde{w}, \\
\eta &= \eta^*\tilde{\eta}, \\
\Phi &= \Phi^*\tilde{\Phi}.
\end{aligned}$$

The non-dimensional variables here represent functions that depend only on other non-dimensional variables. This gives the following, transformed set of equations:

$$\begin{aligned}
\frac{d\tilde{\eta}}{d\tilde{t}} &= -\frac{BT}{\eta^* L} \cdot \frac{\partial \tilde{b}}{\partial \tilde{x}}, \\
\frac{\partial \tilde{\Phi}}{\partial \tilde{x}^2} &= \frac{\eta^*}{\Phi^* (\frac{1}{L^2} + \frac{1}{H^2})} \cdot \tilde{\eta}, \\
\tilde{w} &= -\frac{\Phi^*}{WL} \cdot \frac{\partial \tilde{\Phi}}{\partial \tilde{x}} \text{ and} \\
\tilde{w} &= \frac{H}{WT} \cdot \frac{d\tilde{z}}{d\tilde{t}}.
\end{aligned}$$

The dimensionless factors consisting only of characteristic scales on the right-hand side can be set to equal 1, by determining values for $\eta^*$, $\phi^*$, $T$ and $W$ accordingly:

$$\begin{aligned}
\eta^* &= \frac{BT}{L}, \\
\Phi^* &= \frac{\eta^*}{(\frac{1}{L^2} + \frac{1}{H^2})}, \\
W &= \frac{\Phi^*}{L}, \\
T &= \frac{H}{W}.
\end{aligned}$$

Substituting the first, second and last equation into the third, we obtain the following relationship for $W$:

$$W = \sqrt{\frac{BH}{1 + \frac{L^2}{H^2}}},$$

which describes the characteristic, vertical velocity scale within a circulation driven by horizontal buoyancy gradients and how it depends on both the horizontal and vertical length scales $L$ and $H$ in addition to the characteristic buoyancy scale $B$. See Chapter 5 for further discussion, publications with similar $W$ scalings and the approximation for $\frac{L}{H} \approx 1$. For a detailed introduction to dimensional analysis we refer the reader to Barenblatt (2003) or Witelski and Bowen (2015).

# C

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AR | Autoregressive (process) |
| BRN | Bulk richardson number |
| CAPE | Convectively available potential energy |
| CIN | Convective inhibition |
| COSMO | COnsortium for Small-scale MOdeling |
| CPP | Cold Pool Perturbations |
| DWD | The German weather service (Deutscher Wetterdienst) |
| ECMWF | European center for medium range weather forecasting |
| EDMF | Eddy diffusivity/mass flux |
| FSS | Fraction Skill Score |
| ICON | ICOsahedral Nonhydrostatic [model] |
| ICON-LEM | ICON Large eddy model |
| LES | Large eddy simulations |
| LFC | Level of free convection |
| LNB | Level of neutral buoyancy |
| NWP | Numerical weather prediction |
| PBL | Planetary boundary layer |
| PDF | Probability density function |
| PSP | Physically based stochastic perturbations |
| RDF | Radial Distribution Function |
| SAL | Structure - Amplitude - Location |
| SSO | Subgrid-scale orography |
| SSOSP | SSO stochastic perturbations |

# BIBLIOGRAPHY


Arakawa A, Schubert WH. 1974. Interaction of a Cumulus Cloud Ensemble with the Large-Scale Environment, Part I. *Journal of the Atmospheric Sciences* **31**(3): 674–701, doi: 10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2.

Athey S. 2017. Beyond prediction: Using big data for policy problems. *Science* **355**(6324): 483–485, doi: 10.1126/science.aal4321.

Bachmann K, Keil C, Craig GC, Weissmann M, Welzbacher CA. 2020. Predictability of deep convection in idealized and operational forecasts: Effects of radar data assimilation, orography, and synoptic weather regime. *Monthly Weather Review* **148**(1): 63–81, doi: 10.1175/MWR-D-19-0045.1.

Bachmann K, Keil C, Weissmann M, Branch A. 2019. Impact of Radar Data Assimilation and Orography on Predictability of Deep Convection. *Quarterly Journal of the Royal Meteorological Society* **145**(718): 117–130, doi: 10.1002/qj.3412.

Baines PG, Palmer TN. 1990. Rationale for a new physically-based parametrization of subgrid-scale orographic effects. Technical Report 169, European Centre for Medium Range Weather Forecasting, Shinfield Park, Reading, doi: 10.21957/h4h36b3u.

Baldauf M, Seifert A, Förstner J, Majewski D, Raschendorfer M, Reinhardt T. 2011. Operational convective-scale numerical weather prediction with the COSMO Model: Description and sensitivities. *Monthly Weather Review* **139**(12): 3887–3905, doi: 10.1175/MWR-D-10-05013.1.

Barenblatt GI. 2003. *Scaling*, vol. 34. Cambridge University Press.

Bauer P, Thorpe A, Brunet G. 2015. The quiet revolution of numerical weather prediction. *Nature* **525**(7567): 47, doi: 10.1038/nature14956.

Baur F, Keil C, Craig GC. 2018. Soil moisture–precipitation coupling over Central Europe: Interactions between surface anomalies at different scales and the dynamical implication. *Quarterly Journal of the Royal Meteorological Society* **144**(717): 2863–2875, doi: 10.1002/qj.3415.

Bengtsson L, Steinheimer M, Bechtold P, Geleyn JF. 2013. A stochastic parametrization for deep convection using cellular automata. *Quarterly Journal of the Royal Meteorological Society* **139**(675): 1533–1543, doi: 10.1002/qj.2108.

Benjamin TB. 1968. Gravity currents and related phenomena. *Journal of Fluid Mechanics* **31**(2): 209–248, doi: 10.1017/S0022112068000133.

Berner J, Achatz U, Batté L, Bengtsson L, de la Cámara A, Christensen HM, Colangeli M, Coleman DRB, Crommelin D, Dolaptchiev SI, Franzke CLE, Friederichs P, Imkeller P, Järvinen H, Juricke S, Kitsios V, Lott F, Lucarini V, Mahajan S, Palmer

TN, Penland C, Sakradzija M, von Storch JS, Weisheimer A, Weniger M, Williams PD, Yano JI. 2017. Stochastic Parameterization: Toward a New View of Weather and Climate Models. *Bulletin of the American Meteorological Society* **98**(3): 565–588, doi: 10.1175/BAMS-D-15-00268.1.

Bierdel L, Friederichs P, Bentzien S. 2012. Spatial kinetic energy spectra in the convection-permitting limited-area NWP model COSMO-DE. *Meteorologische Zeitschrift* **21**(3): 245–258, doi: 10.1127/0941-2948/2012/0319.

Böing SJ, Jonker HJJ, Siebesma AP, Grabowski WW. 2012. Influence of the subcloud layer on the development of a deep convective ensemble. *Journal of the Atmospheric Sciences* **69**(9): 2682–2698, doi: 10.1175/JAS-D-11-0317.1.

Brenowitz ND, Bretherton CS. 2018. Prognostic Validation of a Neural Network Unified Physics Parameterization. *Geophysical Research Letters* **45**(12): 6289–6298, doi: 10.1029/2018GL078510.

Brundke F. 2015. Variability in convective initiation through stochastic perturbations representing the effect of subgrid-scale orography. Master's thesis, Ludwig-Maximilians-Universität München.

Bryan GH, Rotunno R. 2008. Gravity Currents in a Deep Anelastic Atmosphere. *Journal of the Atmospheric Sciences* **65**(2): 536–556, doi: 10.1175/2007JAS2443.1.

Bryan GH, Rotunno R. 2014a. Gravity Currents in Confined Channels with Environmental Shear. *Journal of the Atmospheric Sciences* **71**(3): 1121–1142, doi: 10.1175/JAS-D-13-0157.1.

Bryan GH, Rotunno R. 2014b. The Optimal State for Gravity Currents in Shear. *Journal of the Atmospheric Sciences* **71**(1): 448–468, doi: 10.1175/JAS-D-13-0156.1.

Buizza R. 2018. Introduction to the special issue on "25 years of ensemble forecasting". *Quarterly Journal of the Royal Meteorological Society* **145**(S1): 1–11, doi: 10.1002/qj.3370.

Buizza R, Miller M, Palmer T. 1999. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. Technical report, European Centre for Medium-Range Weather Forecasts.

Cafaro C, Rooney GG. 2018. Characteristics of colliding density currents: a numerical and theoretical study. *Quarterly Journal of the Royal Meteorological Society* **144**(715): 1761–1771, doi: 10.1002/qj.3337.

Chagnon JM, Bannon PR. 2005. Adjustment to injections of mass, momentum, and heat in a compressible atmosphere. *Journal of the Atmospheric Sciences* **62**(8): 2749–2769, doi: 10.1175/JAS3503.1.

Chapron B, Dérian P, Mémin E, Resseguier V. 2018. Large-scale flows under location uncertainty: a consistent stochastic framework. *Quarterly Journal of the Royal Meteorological Society* **144**(710), doi: 10.1002/qj.3198.

Chen B, Pearl J. 2015. Graphical tools for linear structural equation modeling. Technical report, California Univ. Los Angeles Dept. of Computer Science, doi: 10.21236/ada609131.

Chow F, Schär C, Ban N, Lundquist K, Schlemmer L, Shi X. 2019. Crossing Multiple Gray Zones in the Transition from Mesoscale to Microscale Simulation over Complex Terrain. *Atmosphere* **10**(5): 274, doi: 10.3390/atmos10050274.

Christensen HM, Lock SJ, Moroz IM, Palmer TN. 2017. Introducing independent patterns into the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme. *Quarterly Journal of the Royal Meteorological Society* **143**(706): 2168–2181, doi: 10.1002/qj.3075.

Clark P, Roberts N, Lean H, Ballard SP, Charlton-Perez C. 2016. Convection-permitting models: a step-change in rainfall forecasting. *Meteorological Applications* **23**(2): 165–181, doi: 10.1002/met.1538.

Craig GC, Cohen BG. 2006. Fluctuations in an Equilibrium Convective Ensemble. Part I: Theoretical Formulation. *Journal of the Atmospheric Sciences* **63**(8): 1996–2004, doi: 10.1175/JAS3709.1.

Craig GC, Selz T. 2018. Mesoscale Dynamical Regimes in the Midlatitudes. *Geophysical Research Letters* doi: 10.1002/2017GL076174.

de Rooy WC, Bechtold P, Fröhlich K, Hohenegger C, Jonker H, Mironov D, Pier Siebesma A, Teixeira J, Yano JI. 2013. Entrainment and detrainment in cumulus convection: an overview. *Quarterly Journal of the Royal Meteorological Society* **139**(670): 1–19, doi: 10.1002/qj.1959.

Demaeyer J, Vannitsem S. 2017. Stochastic parametrization of subgrid-scale processes in coupled ocean–atmosphere systems: benefits and limitations of response theory. *Quarterly Journal of the Royal Meteorological Society* **143**(703): 881–896, doi: 10.1002/qj.2973.

Demaeyer J, Vannitsem S. 2018. *Stochastic parameterization of subgrid-scale processes: A review of recent physically based approaches*. Springer International Publishing: Cham, pp. 55–85, doi: 10.1007/978-3-319-58895-7_3.

Deutscher Wetterdienst. 2018a. Hoch aufgelöste Niederschlgsanalyse und -vorhersage auf der Basis quantitativer Radar- und Omrometerdaten für grenzüberschreitende Fluss-Einzugsgebiete von Deutschland im Echtzeitbetrieb - Beschreibung des Kompositsformats Version 2.4.4. URL `https://www.dwd.de/DE/leistungen/radolan/radolan_info/radolan_radvor_op_komposit_format_pdf.pdf?__blob=publicationFile&v=6`, accessed 2019-13-05.

Deutscher Wetterdienst. 2018b. RADOLAN/RADVOR Produktübersicht. URL `https://www.dwd.de/DE/leistungen/radolan/produktuebersicht/radolan_produktuebersicht_pdf.pdf?__blob=publicationFile&v=6`, accessed 2019-13-05.

Deutscher Wetterdienst. 2019. Hagelunwetter in München und weitere schwere Gewitter an Pfingstmontag . URL https://www.dwd.de/DE/wetter/thema_des_tages/2019/6/11.html, accessed 2020-03-30.

Deutscher Wetterdienst. 2020. Entwicklung des Integrierten Vorhersagesystems SINFONY (Seamless INtegrated FOrecastiNg sYstem) . URL https://www.dwd.de/DE/forschung/forschungsprogramme/sinfony_iafe/sinfony_node.html, accessed 2020-03-25.

Dipankar A, Stevens B, Heinze R, Moseley C, Zängl G, Giorgetta M, Brdar S. 2015. Large eddy simulation using the general circulation model ICON. *Journal of Advances in Modeling Earth Systems* **7**(3): 963–986, doi: 10.1002/2015MS000431.

Doms G, Förstner J, Heise E, Herzog HJ, Mironov D, Raschendorfer M, Reinhardt T, Ritter B, Schrodin R, Schulz JP, Vogel G. 2011. A description of the nonhydrostatic regional COSMO model. Part II: physical parameterization. *Deutscher Wetterdienst, Offenbach, Germany* .

Drager AJ, van den Heever SC. 2017. Characterizing convective cold pools. *Journal of Advances in Modeling Earth Systems* **9**(2): 1091–1115, doi: 10.1002/2016MS000788.

Ebert-Uphoff I, Deng Y. 2012. Causal discovery for climate research using graphical models. *Journal of Climate* **25**(17): 5648–5665, doi: 10.1175/JCLI-D-11-00387.10.

ECMWF. 2016. The strength of a common goal - A roadmap to 2025. Technical report, accessed 2019-11-07, URL https://www.ecmwf.int/sites/default/files/ECMWF_Roadmap_to_2025.pdf.

ECMWF. 2019. Operational configurations of the ECMWF Integrated Forecasting System (IFS). URL https://www.ecmwf.int/en/forecasts/documentation-and-support, accessed 2019-11-07.

Edson AR, Bannon PR. 2008. Nonlinear Atmospheric Adjustment to Thermal Forcing. *Journal of the Atmospheric Sciences* **62**(12): 4253–4272, doi: 10.1175/JAS3517.1.

Ehrendorfer M. 1994a. The Liouville Equation and Its Potential Usefulness for the Prediction of Forecast Skill. Part I: Theory. *Monthly Weather Review* **122**(4): 703–713, doi: 10.1175/1520-0493(1994)122<0703:TLEAIP>2.0.CO;2.

Ehrendorfer M. 1994b. The Liouville Equation and Its Potential Usefulness for the Prediction of Forecast Skill. Part II: Applications. *Monthly Weather Review* **122**(4): 714–728, doi: 10.1175/1520-0493(1994)122<0714:TLEAIP>2.0.CO;2.

Emanuel KA. 1994. *Atmospheric convection*. Oxford University Press.

Feng Z, Hagos S, Rowe AK, Burleyson CD, Martini MN, de Szoeke SP. 2015. Mechanisms of convective cloud organization by cold pools over tropical warm ocean during the AMIE/DYNAMO field campaign. *Journal of Advances in Modeling Earth Systems* **7**(2): 357–381, doi: 10.1002/2014MS000384.

Fiedler BH. 2002. A wind transform for acoustic adjustment in compressible models. *Monthly Weather Review* **130**(3): 741–746, doi: 10.1175/1520-0493(2002)130<0741:AWTFAA>2.0.CO;2.

Fournier MB, Haerter JO. 2019. Tracking the gust fronts of convective cold pools. *Journal of Geophysical Research: Atmospheres* **124**(21): 11 103–11 117, doi: 10.1029/2019JD030980.

Franzke CLE, O'Kane TJ, Berner J, Williams PD, Lucarini V. 2015. Stochastic climate theory and modeling. *Wiley Interdisciplinary Reviews: Climate Change* **6**(1): 63–78, doi: 10.1002/wcc.318.

Geerts B, Miao Q, Demko JC. 2008. Pressure Perturbations and Upslope Flow over a Heated, Isolated Mountain. *Monthly Weather Review* **136**(11): 4272–4288, doi: 10.1175/2008MWR2546.1.

Gentine P, Pritchard M, Rasp S, Reinaudi G, Yacalis G. 2018. Could Machine Learning Break the Convection Parameterization Deadlock? *Geophysical Research Letters* **45**(11): 5742–5751, doi: 10.1029/2018GL078202.

Gill AE. 1982. *Atmosphere-ocean dynamics*. Academic Press, ISBN 9780080570525.

Gilleland E, Ahijevych D, Brown BG, Casati B, Ebert EE. 2009. Intercomparison of Spatial Forecast Verification Methods. *Weather and Forecasting* **24**(5): 1416–1430, doi: 10.1175/2009WAF2222269.1.

Golding B, Mylne K, Clark P. 2004. The history and future of numerical weather prediction in the Met Office. *Weather* **59**(11): 299–306, doi: 10.1256/wea.113.04.

Grabowski WW. 2004. An Improved Framework for Superparameterization. *Journal of the Atmospheric Sciences* **61**(15): 1940–1952, doi: 10.1175/1520-0469(2004)061<1940:AIFFS>2.0.CO;2.

Grandpeix JY, Lafore JP. 2010. A density current parameterization coupled with Emanuel's convection scheme. Part I: The models. *Journal of the Atmospheric Sciences* **67**(4): 881–897, doi: 10.1175/2009JAS3044.1.

Grandpeix JY, Lafore JP, Cheruy F. 2010. A density current parameterization coupled with Emanuel's convection scheme. Part II: 1D simulations. *Journal of the Atmospheric Sciences* **67**(4): 898–922, doi: 10.1175/2009JAS3045.1.

Grant LD, van den Heever SC. 2016. Cold pool dissipation. *Journal of Geophysical Research: Atmospheres* **121**(3): 1138–1155.

Grant LD, van den Heever SC. 2018. Cold Pool-Land Surface Interactions in a Dry Continental Environment. *Journal of Advances in Modeling Earth Systems* **10**(7): 1513–1526, doi: 10.1029/2018MS001323.

Gustafsson N, Janjić T, Schraff C, Leuenberger D, Weissmann M, Reich H, Brousseau P, Montmerle T, Wattrelot E, Bučánek A, Mile M, Hamdi R, Lindskog M, Barkmeijer J, Dahlbom M, Macpherson B, Ballard S, Inverarity G, Carley J, Alexander C, Dowell D,

Liu S, Ikuta Y, Fujita T. 2018. Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society* **144**(713): 1218–1256, doi: 10.1002/qj.3179.

Haerter JO, Boeing SJ, Henneberg O, Nissen SB. 2018. Reconciling cold pool dynamics with convective self-organization. *arXiv preprint; arXiv:1810.05518* .

Hanley K, Whitall M, Stirling A, Clark P. 2019. Modifications to the representation of subgrid mixing in kilometre-scale versions of the Unified Model. *Quarterly Journal of the Royal Meteorological Society* **145**(725): 3361–3375, doi: 10.1002/qj.3624.

Hanley KE, Plant RS, Stein THM, Hogan RJ, Nicol JC, Lean HW, Halliwell C, Clark PA. 2015. Mixing-length controls on high-resolution simulations of convective storms. *Quarterly Journal of the Royal Meteorological Society* **141**(686): 272–284.

Harper K, Uccellini LW, Kalnay E, Carey K, Morone L. 2007. Symposium of the 50th anniversary of operational numerical weather prediction. *Bulletin of the American Meteorological Society* **88**(5): 639–650, doi: 10.1175/BAMS-88-5-639.

Heinze R, Dipankar A, Henken CC, Moseley C, Sourdeval O, Trömel S, Xie X, Adamidis P, Ament F, Baars H, Barthlott C, Behrendt A, Blahak U, Bley S, Brdar S, Brueck M, Crewell S, Deneke H, Di Girolamo P, Evaristo R, Fischer J, Frank C, Friederichs P, Göcke T, Gorges K, Hande L, Hanke M, Hansen A, Hege HC, Hoose C, Jahns T, Kalthoff N, Klocke D, Kneifel S, Knippertz P, Kuhn A, van Laar T, Macke A, Maurer V, Mayer B, Meyer CI, Muppa SK, Neggers RAJ, Orlandi E, Pantillon F, Pospichal B, Röber N, Scheck L, Seifert A, Seifert P, Senf F, Siligam P, Simmer C, Steinke S, Stevens B, Wapler K, Weniger M, Wulfmeyer V, Zängl G, Zhang D, Quaas J. 2017. Large-eddy simulations over Germany using ICON: a comprehensive evaluation. *Quarterly Journal of the Royal Meteorological Society* **143**(702): 69–100, doi: 10.1002/qj.2947.

Henneberg O, Meyer B, Haerter JO. 2020. Particle-based tracking of cold pool gust fronts. *Journal of Advances in Modeling Earth Systems* : e2019MS001 910doi: 10.1029/2019MS001910.

Hohenegger C, Schär C. 2007. Atmospheric predictability at synoptic versus cloud-resolving scales. *Bulletin of the American Meteorological Society* **88**(11): 1783–1793, doi: 10.1175/BAMS-88-11-1783.

Honnert R. 2016. Representation of the grey zone of turbulence in the atmospheric boundary layer. *Advances in Science and Research* **13**: 63–67, doi: 10.5194/asr-13-63-2016.

Hourdin F, Mauritsen T, Gettelman A, Golaz JC, Balaji V, Duan Q, Folini D, Ji D, Klocke D, Qian Y, Rauser F, Rio C, Tomassini L, Watanabe M, Williamson D. 2017. The art and science of climate model tuning. *Bulletin of the American Meteorological Society* **98**(3): 589–602, doi: 10.1175/BAMS-D-15-00135.1.

Houze RA. 2012. Orographic effects on precipitating clouds. *Reviews of Geophysics* **50**(1): RG1001, doi: 10.1029/2011RG000365.

Igel AL, Igel MR, van den Heever SC. 2015. Make It a Double? Sobering Results from Simulations Using Single-Moment Microphysics Schemes. *Journal of the Atmospheric Sciences* **72**(2): 910–925, doi: 10.1175/JAS-D-14-0107.1.

Jankov I, Beck J, Wolff J, Harrold M, Olson JB, Smirnova T, Alexander C, Berner J. 2019. Stochastically Perturbed Parameterizations in an HRRR-Based Ensemble. *Monthly Weather Review* **147**(1): 153–173, doi: 10.1175/MWR-D-18-0092.1.

Jankov I, Berner J, Beck J, Jiang H, Olson JB, Grell G, Smirnova TG, Benjamin SG, Brown JM. 2017. A performance comparison between multiphysics and stochastic approaches within a North American RAP ensemble. *Monthly Weather Review* **145**(4): 1161–1179.

Jeevanjee N. 2017. Vertical Velocity in the Gray Zone. *Journal of Advances in Modeling Earth Systems* **9**(6): 2304–2316, doi: 10.1002/2017MS001059.

Jeevanjee N, Romps DM. 2016. Effective buoyancy at the surface and aloft. *Quarterly Journal of the Royal Meteorological Society* **142**(695): 811–820, doi: 10.1002/qj.2683.

Judt F. 2018. Insights into atmospheric predictability through global convection-permitting model simulations. *Journal of the Atmospheric Sciences* **75**(5): 1477–1497, doi: 10.1175/JAS-D-17-0343.1.

Kealy JC. 2019. Probing the 'grey zone' of NWP – is higher resolution always better? *Weather* **74**(7): 246–249, doi: 10.1002/wea.3506.

Kealy JC, Efstathiou GA, Beare RJ. 2019. The onset of resolved boundary-layer turbulence at grey-zone resolutions. *Boundary-Layer Meteorology* **171**(1): 31–52, doi: 10.1007/s10546-018-0420-0.

Keil C, Baur F, Bachmann K, Rasp S, Schneider L, Barthlott C. 2019. Relative contribution of soil moisture, boundary-layer and microphysical perturbations on convective predictability in different weather regimes. *Quarterly Journal of the Royal Meteorological Society* **145**(724): 3102–3115, doi: 10.1002/qj.3607.

Keil C, Heinlein F, Craig GC. 2014. The convective adjustment time-scale as indicator of predictability of convective precipitation. *Quarterly Journal of the Royal Meteorological Society* **140**(679): 480–490, doi: 10.1002/qj.2143.

Kirshbaum D, Adler B, Kalthoff N, Barthlott C, Serafin S. 2018. Moist orographic convection: Physical mechanisms and links to surface-exchange processes. *Atmosphere* **9**(3): 80, doi: 10.3390/atmos9030080.

Kirshbaum DJ, Rotunno R, Bryan GH. 2007a. The spacing of orographic rainbands triggered by small-scale topography. *Journal of the Atmospheric Sciences* **64**(12): 4222–4245, doi: 10.1175/2007JAS2335.1.

Kirshbaum DJ, Rotunno R, Bryan GH, Durran DR. 2007b. The Triggering of Orographic Rainbands by Small-Scale Topography. *Journal of the Atmospheric Sciences* **64**(12): 4222–4245, doi: 10.1175/2007JAS2335.1.

Kober K, Craig GC. 2016. Physically Based Stochastic Perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *Journal of the Atmospheric Sciences* **73**(7): 2893–2911, doi: 10.1175/JAS-D-15-0144.1.

Kretschmer M, Coumou D, Donges JF, Runge J. 2016. Using causal effect networks to analyze different Arctic drivers of midlatitude winter circulation. *Journal of Climate* **29**(11): 4069–4081, doi: 10.1175/JCLI-D-15-0654.1.

Krich C, Runge J, Miralles DG, Migliavacca M, Perez-Priego O, El-Madany T, Carrara A, Mahecha MD. 2020. Estimating causal networks in biosphere–atmosphere interaction with the pcmci approach. *Biogeosciences* **17**(4): 1033–1061, doi: 10.5194/bg-17-1033-2020.

Kron W, Eichner J, Kundzewicz ZW. 2019. Reduction of flood risk in Europe – Reflections from a reinsurance perspective. *Journal of Hydrology* **576**: 197–209, doi: 10.1016/j.jhydrol.2019.06.050.

Langhans W, Gohm A, Zängl G. 2011. The orographic impact on patterns of embedded convection during the August 2005 Alpine flood. *Quarterly Journal of the Royal Meteorological Society* **137**(661): 2092–2105.

Langhans W, Schmidli J, Fuhrer O, Bieri S, Schar C. 2013. Long-term simulations of thermally driven flows and orographic convection at convection-parameterizing and cloud-resolving resolutions. *Journal of Applied Meteorology and Climatology* **52**(6): 1490–1510, doi: 10.1175/JAMC-D-12-0167.1.

Leutwyler D, Lüthi D, Ban N, Fuhrer O, Schär C. 2017. Evaluation of the convection-resolving climate modeling approach on continental scales. *Journal of Geophysical Research: Atmospheres* **122**(10): 5237–5258, doi: 10.1002/2016JD026013.

Lilly DK. 1990. Numerical prediction of thunderstorms—has its time come? *Quarterly Journal of the Royal Meteorological Society* **116**(494): 779–798, doi: 10.1002/qj.49711649402.

Lott F, Miller MJ. 1997. A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Quarterly Journal of the Royal Meteorological Society* **123**: 101–127, doi: 10.1002/qj.49712353704.

Markowski P, Richardson Y. 2011. *Mesoscale meteorology in midlatitudes*, vol. 2. John Wiley & Sons.

Mauritsen T, Stevens B, Roeckner E, Crueger T, Esch M, Giorgetta M, Haak H, Jungclaus J, Klocke D, Matei D, Mikolajewicz U, Notz D, Pincus R, Schmidt H, Tomassini L. 2012. Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems* **4**(3): 1–18, doi: 10.1029/2012MS000154.

Mayer B. 2018. Erwärmungs- und Abkühlungsraten - Wie wichtig ist der 3D-Strahlungsreport? *promet* **100**: 98–110.

Mellor GL, Yamada T. 1982. Development of a turbulence closure model for geophysical fluid problems. *Reviews of Geophysics* **20**(4): 851, doi: 10.1029/RG020i004p00851.

Moncrieff MW, Liu C. 1999. Convection Initiation by Density Currents: Role of Convergence, Shear, and Dynamical Organization. *Monthly Weather Review* **127**(10): 2455–2464, doi: 10.1175/1520-0493(1999)127<2455:CIBDCR>2.0.CO;2.

Morrison H. 2016a. Impacts of updraft size and dimensionality on the perturbation pressure and vertical velocity in cumulus convection. Part I: Simple, generalized analytic solutions. *Journal of the Atmospheric Sciences* **73**(4): 1441–1454, doi: 10.1175/JAS-D-15-0040.1.

Morrison H. 2016b. Impacts of updraft size and dimensionality on the perturbation pressure and vertical velocity in cumulus convection. Part II: Comparison of theoretical and numerical solutions and fully dynamical simulations. *Journal of the Atmospheric Sciences* **73**(4): 1455–1480, doi: 10.1175/JAS-D-15-0041.1.

Morrison H, Morales A, Villanueva-Birriel C. 2015. Concurrent sensitivities of an idealized deep convective storm to parameterization of microphysics, horizontal grid resolution, and environmental static stability. *Monthly Weather Review* **143**(6): 2082–2104, doi: 10.1175/MWR-D-14-00271.1.

Munich Re. 2020. Wirbelstürme mit Milliardenschäden prägen Bilanz der Naturkatastrophen 2019 . URL `https://www.munichre.com/de/unternehmen/media-relations/medieninformationen-und-unternehmensnachrichten/medieninformationen/2020/milliardenschaeden-praegen-bilanz-naturkatastrophen-2019.html`, accessed 2020-03-30.

Nebeker F. 1997. Calculating the Weather: Meteorology in the 20th Century. *Technology and Culture* **38**(3): 786, doi: 10.2307/3106892.

Necker T, Geiss S, Weissmann M, Ruiz J, Miyoshi T, Lien GY. 2020. A convective-scale 1,000-member ensemble simulation and potential applications. *Quarterly Journal of the Royal Meteorological Society* doi: 10.1002/qj.3744.

Necker T, Weissmann M, Sommer M. 2018. The importance of appropriate verification metrics for the assessment of observation impact in a convection-permitting modelling system. *Quarterly Journal of the Royal Meteorological Society* **144**(714): 1667–1680, doi: 10.1002/qj.3390.

Nowack P, Runge J, Eyring V, Haigh JD. 2020. Causal networks for climate model evaluation and constrained projections. *Nature Communications* **11**(1): 1–11, doi: 10.1038/s41467-020-15195-y.

Ollinaho P, Lock SJ, Leutbecher M, Bechtold P, Beljaars A, Bozzo A, Forbes RM, Haiden T, Hogan RJ, Sandu I. 2017. Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble. *Quarterly Journal of the Royal Meteorological Society* **143**(702): 408–422, doi: 10.1002/qj.2931.

Palmer T. 2019a. The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years. *Quarterly Journal of the Royal Meteorological Society* **145**(S1): 12–24, doi: 10.1002/qj.3383.

Palmer TN. 2019b. Stochastic weather and climate models. *Nature Reviews Physics* **1**(7): 463–471, doi: 10.1038/s42254-019-0062-2.

Palmer TN, Weisheimer A. 2011. Diagnosing the causes of bias in climate models–why is it so hard? *Geophysical & Astrophysical Fluid Dynamics* **105**(2-3): 351–365, doi: 10.1080/03091929.2010.547194.

Panosetti D, Böing S, Schlemmer L, Schmidli J. 2016. Idealized large-eddy and convection-resolving simulations of moist convection over mountainous terrain. *Journal of the Atmospheric Sciences* **73**(10): 4021–4041, doi: 10.1175/JAS-D-15-0341.1.

Panosetti D, Schlemmer L, Schär C. 2019. Bulk and structural convergence at convection-resolving scales in real-case simulations of summertime moist convection over land. *Quarterly Journal of the Royal Meteorological Society* **145**(721): 1427–1443, doi: 10.1002/qj.3502.

ParaCon. 2020. ParaCon - Representation of convection in models - Met Office. URL `https://www.metoffice.gov.uk/research/approach/collaboration/paracon`, accessed 2020-01-29.

Parishani H, Pritchard MS, Bretherton CS, Wyant MC, Khairoutdinov M. 2017. Toward low-cloud-permitting cloud superparameterization with explicit boundary layer turbulence. *Journal of Advances in Modeling Earth Systems* **9**(3): 1542–1571, doi: 10.1002/2017MS000968.

Park S. 2014. A Unified Convection Scheme (UNICON). Part I: Formulation. *Journal of the Atmospheric Sciences* **71**(11): 3902–3930, doi: 10.1175/JAS-D-13-0233.1.

Pauluis O, Garner S. 2006. Sensitivity of Radiative–Convective Equilibrium Simulations to Horizontal Resolution. *Journal of the Atmospheric Sciences* **63**(7): 1910–1923, doi: 10.1175/JAS3705.1.

Pearl J. 2009. *Causality*. Cambridge university press.

Pearl J. 2013. Linear models: A useful "microscope" for causal analysis. Technical report, University of California, Los Angeles Computer Science Department.

Pearl J. 2014. Interpretation and identification of causal mediation. *Psychological methods* **19**(4): 459, doi: 10.1037/a0036434.

Pearl J. 2019. The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM* **62**(3): 54–60, doi: 10.1145/3241036.

Penland C. 2003. Noise out of chaos and why it won't go away. *Bulletin of the American Meteorological Society* **84**(7): 921–926, doi: 10.1175/BAMS-84-7-921.

Peters K, Hohenegger C. 2017. On the dependence of squall-line characteristics on surface conditions. *Journal of the Atmospheric Sciences* **74**(7): 2211–2228, doi: 10.1175/JAS-D-16-0290.1.

Piper D, Kunz M, Ehmele F, Mohr S, Mühr B, Kron A, Daniell J. 2016. Exceptional sequence of severe thunderstorms and related flash floods in May and June 2016 in Germany – Part 1: Meteorological background. *Natural Hazards and Earth System Sciences* **16**(12): 2835–2850, doi: 10.5194/nhess-16-2835-2016.

Plant RS, Craig GC. 2008. A stochastic parameterization for deep convection based on equilibrium statistics. *Journal of the Atmospheric Sciences* **65**(1): 87–105, doi: 10.1175/2007JAS2263.1.

Pscheidt I, Senf F, Heinze R, Deneke H, Trömel S, Hohenegger C. 2019. How organized is deep convection over Germany? *Quarterly Journal of the Royal Meteorological Society* **145**(723): 2366–2384, doi: 10.1002/qj.3552.

Randall DA, Bitz CM, Danabasoglu G, Denning AS, Gent PR, Gettelman A, Griffies SM, Lynch P, Morrison H, Pincus R, Thuburn J. 2018. 100 Years of Earth System Model Development. *Meteorological Monographs* **59**: 12.1–12.66, doi: 10.1175/AMSMONOGRAPHS-D-18-0018.1.

Raschendorfer M. 2001. The new turbulence parameterization of LM. *COSMO newsletter* **1**: 89–97.

Rasp S, Pritchard MS, Gentine P. 2018a. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences* **115**(39): 9684–9689, doi: 10.1073/pnas.1810286115.

Rasp S, Selz T, Craig GC. 2018b. Variability and Clustering of Midlatitude Summertime Convection: Testing the Craig and Cohen Theory in a Convection-Permitting Ensemble with Stochastic Boundary Layer Perturbations. *Journal of the Atmospheric Sciences* **75**(2): 691–706, doi: 10.1175/JAS-D-17-0258.1.

Reinert D, Prill F, Frank H, Denhard M, Zängl G. 2018. Database Reference Manual for ICON and ICON-EPS. Technical Report Version 1.2.11, Deutscher Wetterdienst, Offenbach am Main, doi: 10.5676/DWD_pub/nwv/icon_1.2.11.

Rio C, Del Genio AD, Hourdin F. 2019. Ongoing Breakthroughs in Convective Parameterization. *Current Climate Change Reports* **5**(2): 95–111, doi: 10.1007/s40641-019-00127-w.

Roberts NM, Lean HW. 2008. Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Monthly Weather Review* **136**(1): 78–97, doi: 10.1175/2007MWR2123.1.

Rotach MW, Gohm A, Lang MN, Leukauf D, Stiperski I, Wagner JS. 2015. On the Vertical Exchange of Heat, Mass, and Momentum Over Complex, Mountainous Terrain. *Frontiers in Earth Science* **3**(43): 514–515, doi: 10.3389/feart.2015.00076.

Rotunno R, Klemp JB, Weisman ML. 1988. A theory for strong, long-lived squall lines. *Journal of the Atmospheric Sciences* **45**(3): 463–485, doi: 10.1175/1520-0469(1988)045<0463:ATFSLL>2.0.CO;2.

Rozbicki JJ, Young GS, Qian L. 1999. Test of a Convective Wake Parameterization in the Single-Column Version of CCM3. *Monthly Weather Review* **127**(6): 1347–1361, doi: 10.1175/1520-0493(1999)127<1347:TOACWP>2.0.CO;2.

Ruckstuhl Y, Janjić T. 2020. Combined state-parameter estimation with the letkf for convective-scale weather forecasting. *Monthly Weather Review* **148**(4): 1607–1628, doi: 10.1175/MWR-D-19-0233.1.

Runge J, Bathiany S, Bollt E, Camps-Valls G, Coumou D, Deyle E, Glymour C, Kretschmer M, Mahecha MD, Muñoz-Marí J, van Nes EH, Peters J, Quax R, Reichstein M, Scheffer M, Schölkopf B, Spirtes P, Sugihara G, Sun J, Zhang K, Zscheischler J. 2019. Inferring causation from time series in Earth system sciences. *Nature Communications* **10**(1): 2553, doi: 10.1038/s41467-019-10105-3.

Runge J, Petoukhov V, Donges JF, Hlinka J, Jajcay N, Vejmelka M, Hartman D, Marwan N, Paluš M, Kurths J. 2015. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature communications* **6**: 8502, doi: 10.1038/ncomms9502.

Sakradzija M, Seifert A, Dipankar A. 2016. A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone. *Journal of Advances in Modeling Earth Systems* **8**(2): 786–812, doi: 10.1002/2016MS000634.

Samarasinghe SM, McGraw MC, Barnes EA, Ebert-Uphoff I. 2019. A study of links between the Arctic and the midlatitude jet stream using Granger and Pearl causality. *Environmetrics* **30**(4): e2540, doi: 10.1002/env.2540.

Schär C, Fuhrer O, Arteaga A, Ban N, Charpilloz C, Di Girolamo S, Hentgen L, Hoefler T, Lapillonne X, Leutwyler D, Osterried K, Panosetti D, Rüdisühli S, Schlemmer L, Schulthess T, Sprenger M, Ubbiali S, Wernli H. 2019. Kilometer-scale climate models: Prospects and challenges. *Bulletin of the American Meteorological Society* doi: 10.1175/BAMS-D-18-0167.1.

Schättler U, Doms G, Schraff C. 2016. A description of the nonhydrostatic regional COSMO model, Part VII: User's guide, Consortium for Small-Scale Modelling (COSMO). *Deutscher Wetterdienst, Offenbach, Germany* .

Schlemmer L, Hohenegger C. 2014. The formation of wider and deeper clouds as a result of cold-pool dynamics. *Journal of the Atmospheric Sciences* **71**(8): 2842–2858, doi: 10.1175/JAS-D-13-0170.1.

Schölkopf B. 2019. Causality for machine learning. *arXiv preprint; arXiv:1911.10500v2* .

Schneider L, Barthlott C, Barrett AI, Hoose C. 2018. The precipitation response to variable terrain forcing over low mountain ranges in different weather regimes. *Quarterly Journal of the Royal Meteorological Society* **144**(713): 970–989, doi: 10.1002/qj.3250.

Schraff C, Reich H, Rhodin A, Schomburg A, Stephan K, Periáñez A, Potthast R. 2016. Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quarterly Journal of the Royal Meteorological Society* **142**(696): 1453–1472, doi: 10.1002/qj.2748.

Seifert A, Beheng KD. 2006. A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description. *Meteorology and Atmospheric Physics* **92**(1-2): 45–66, doi: 10.1007/s00703-005-0112-4.

Senf F, Klocke D, Brueck M. 2018. Size-Resolved Evaluation of Simulated Deep Tropical Convection. *Monthly Weather review* **146**(7): 2161–2182, doi: 10.1175/MWR-D-17-0378.1.

Shutts G. 2005. A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society* **131**(612): 3079–3102, doi: 10.1256/qj.04.106.

Siebesma AP, Soares PMM, Teixeira J. 2007. A Combined Eddy-Diffusivity Mass-Flux Approach for the Convective Boundary Layer. *Journal of the Atmospheric Sciences* **64**(4): 1230–1248, doi: 10.1175/JAS3888.1.

Squitieri BJ, Gallus WA. 2020. On the forecast sensitivity of mcs cold pools and related features to horizontal grid spacing in convection-allowing wrf simulations. *Weather and Forecasting* **35**(2): 325–346, doi: 10.1175/WAF-D-19-0016.1.

Stensrud DJ. 2007. *Parameterization Schemes*, vol. 9780521865. Cambridge University Press: Cambridge, ISBN 9780511812590.

Stevens B, Acquistapace C, Hansen A, Heinze R, Klinger C, Klocke D, Schubotz W, Windmiller J, Team HAV. 2019a. Large-eddy and Storm Resolving Models for Climate Prediction - The Added Value for Clouds and Precipitation. *Journal of the Meteorological Society Japan* **in review**.

Stevens B, Satoh M, Auger L, Biercamp J, Bretherton C, Düben P, Judt F, Khairoutdinov M, Klocke D, Kornblueh L, Kodama C, Neumann P, Lin S, Putman WM, Röber N, Shibuya R, Vidale P, Wedi N. 2019b. DYAMOND: The DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Prog. In Earth and Planet. Sci* : 1–18.

Stull RB. 1988. *An introduction to boundary layer meteorology*. Kluwer Academic Publishers, ISBN 9027727686.

Tan Z, Kaul CM, Pressel KG, Cohen Y, Schneider T, Teixeira J. 2018. An Extended Eddy-Diffusivity Mass-Flux Scheme for Unified Representation of Subgrid-Scale Turbulence and Convection. *Journal of Advances in Modeling Earth Systems* **10**(3): 770–800, doi: 10.1002/2017MS001162.

Tang SL, Kirshbaum DJ. 2020. On the sensitivity of deep-convection initiation to horizontal grid resolution. *Quarterly Journal of the Royal Meteorological Society* doi: 10.1002/qj.3726.

Thuburn J, Efstathiou GA, Beare RJ. 2019. A two-fluid single-column model of the dry, shear-free, convective boundary layer. *Quarterly Journal of the Royal Meteorological Society* **145**(721): 1535–1550, doi: 10.1002/qj.3510.

Thuburn J, Weller H, Vallis GK, Beare RJ, Whitall M. 2018. A Framework for convection and boundary layer parameterization derived from conditional filtering. *Journal of the Atmospheric Sciences* **75**(3): 965–981, doi: 10.1175/JAS-D-17-0130.1.

Tomassini L, Honnert R, Efstathiou G, Lock A, Siebesma AP. 2019. Second phase of the Grey Zone Project. URL `https://www.metoffice.gov.uk/research/approach/collaboration/grey-zone-project/index`, accessed 2019-09-24.

Tompkins AM. 2001. Organization of Tropical Convection in Low Vertical Wind Shears: The Role of Water Vapor. *Journal of the Atmospheric Sciences* **58**(6): 529–545, doi: 10.1175/1520-0469(2001)058<1650:OOTCIL>2.0.CO;2.

Torri G, Kuang Z. 2019. On Cold Pool Collisions in Tropical Boundary Layers. *Geophysical Research Letters* **46**(1): 399–407, doi: 10.1029/2018GL080501.

Torri G, Kuang Z, Tian Y. 2015. Mechanisms for convection triggering by cold pools. *Geophysical Research Letters* **42**(6): 1943–1950, doi: 10.1002/2015GL063227.

Trapp RJ. 2013. *Mesoscale-convective processes in the atmosphere*. Cambridge University Press.

Tucker DF, Crook NA. 2005. Flow over heated terrain. Part II: Generation of convective precipitation'. *Monthly Weather Review* **133**(9): 2565–2582, doi: 10.1175/MWR2965.1.

Turcotte DL. 1987. A fractal interpretation of topography and geoid spectra on the Earth, Moon, Venus, and Mars. *Journal of Geophysical Research: Solid Earth* **92**(B4): E597–E601, doi: 10.1029/JB092iB04p0E597.

Vallis GK. 2017. *Atmospheric and oceanic fluid dynamics: Fundamentals and large-scale circulation, second edition*. Cambridge University Press: Cambridge, ISBN 9781107588417.

Vissio G, Lucarini V. 2018. A proof of concept for scale-adaptive parametrizations: the case of the Lorenz '96 model. *Quarterly Journal of the Royal Meteorological Society* **144**(710): 63–75, doi: 10.1002/qj.3184.

Vogelezang DHP, Holtslag AAM. 1996. Evaluation and model impacts of alternative boundary-layer height formulations. *Boundary-Layer Meteorology* **81**(3-4): 245–269, doi: 10.1007/BF02430331.

von Kármán T. 1940. The engineer grapples with nonlinear problems. *Bulletin of the American Mathematical Society* **46**(8): 615–684, doi: 10.1090/S0002-9904-1940-07266-0.

Wagner JS, Gohm A, Rotach MW. 2014. The Impact of Horizontal Model Grid Resolution on the Boundary Layer Structure over an Idealized Valley. *Monthly Weather Review* **142**(9): 3446–3465, doi: 10.1175/MWR-D-14-00002.1.

Warner TT. 2010. *Numerical Weather and Climate Prediction*, vol. 9780521513. Cambridge University Press: Cambridge, ISBN 9780511763243.

Weisman ML, Rotunno R. 2004. "A Theory for Strong Long-Lived Squall Lines" Revisited. *Journal of the Atmospheric Sciences* **61**(4): 361–382, doi: 10.1175/1520-0469(2004)061<0361:ATFSLS>2.0.CO;2.

Weisman ML, Skamarock WC, Klemp JB. 1997. The resolution dependence of explicitly modeled convective systems. *Monthly Weather Review* **125**(4): 527–548, doi: 10.1175/1520-0493(1997)125<0527:TRDOEM>2.0.CO;2.

Wernli H, Hofmann C, Zimmer M. 2009. Spatial forecast verification methods intercomparison project: application of the SAL technique. *Weather and Forecasting* **24**(6): 1472–1484, doi: 10.1175/2009WAF2222271.1.

Wernli H, Paulat M, Hagen M, Frei C. 2008. SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Monthly Weather Review* **136**: 4470–4487, doi: 10.1175/2008MWR2415.1.

Wetzel PJ. 1982. Toward parameterization of the stable boundary layer. *Journal of Applied Meteorology* **21**(1): 7–13, doi: 10.1175/1520-0450(1982)021<0007:TPOTSB>2.0.CO;2.

Wilks DS. 2005. *Statistical Methods in the Atmospheric Sciences*. Academic Press, ISBN 0080456227.

Witelski T, Bowen M. 2015. *Methods of mathematical modelling*. Springer, ISBN 978-3-319-23042-9.

Wright S. 1921. Correlation and causation. *Journal of Agricultural Research* **20**: 557–585.

Wyngaard JC. 2004. Toward numerical modeling in the "Terra Incognita". *Journal of the Atmospheric Sciences* **61**(14): 1816–1826, doi: 10.1175/1520-0469(2004)061<1816:TNMITT>2.0.CO;2.

Wyngaard JC. 2010. *Turbulence in the Atmosphere*. Cambridge University Press: Cambridge, ISBN 9780511840524, doi: 10.1017/CBO9780511840524.

Yano JI. 2016. Subgrid-scale physical parameterization in atmospheric modeling: How can we make it consistent? *Journal of Physics A: Mathematical and Theoretical* **49**(28): 284 001, doi: 10.1088/1751-8113/49/28/284001.

Yano JI, Ziemiański MZ, Cullen M, Termonia P, Onvlee J, Bengtsson L, Carrassi A, Davy R, Deluca A, Gray SL, Homar V, Köhler M, Krichak S, Michaelides S, Phillips VTJ, Soares PMM, Wyszogrodzki AA. 2018. Scientific Challenges of Convective-Scale Numerical Weather Prediction. *Bulletin of the American Meteorological Society* **99**(4): 699–710, doi: 10.1175/BAMS-D-17-0125.1.

Zängl G, Reinert D, Rípodas P, Baldauf M. 2015. The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society* **141**(687): 563–579, doi: 10.1002/qj.2378.

Zeng Y, Janjić T, De Lozar A, Rasp S. 2020. Comparison of methods accounting for subgrid-scale model error in convective-scale data assimilation. *Monthly Weather Review* **accepted**, doi: 10.1175/MWR-D-19-0064.1.

Zhou L, Lin SJ, Chen JH, Harris LM, Chen X, Rees SL. 2019. Toward convective-scale prediction within the next generation global prediction system. *Bulletin of the American Meteorological Society* **100**(7): 1225–1243, doi: 10.1175/bams-d-17-0246.1.

# ACKNOWLEDGEMENTS

First and foremost, I am extremely grateful for my supervisor George Craig. Thank you for providing me with an exciting and challenging research topic for my Ph.D. Even more importantly, I am grateful for your very constructive feedback, endless ideas and invaluable, motivating guidance during these last three years. I am looking forward to continuing our joint research on a fascinating new topic.

Thank you, Mark Wenig, for co-examining my thesis.

Big thanks also go to Stephan and Kevin. I highly appreciated having two, more experienced Ph.D. students by my side who regularly checked in on me and advised me on bigger and smaller day-to-day problems. You certainly helped me to better organize myself, to improve my coding skills and to become a better scientist. I miss working and bouldering with you!

Thank you, Yvonne, Florian and Matthias (and many others), for helping out especially when Stephan and Kevin left!

Thank you, Robert, for magically fixing technical problems, for making sure everything runs smoothly or for improving my workflow with ensemble tools or jupyterhub.

Furthermore, I would like to thank the whole LS-Craig team, especially Julien, Leo, Tobi N., Tobi S., Federico and many more, for helping with technical problems, giving me feedback on my thesis or helping with other scientific challenges. I also very much enjoyed our invaluable lunch, party-cracker or ice cream conversations. Home office just cannot compete with that.

I also highly appreciate the Waves-to-Weather project and its community. Certainly, its funding crucially contributed to this thesis, but the W2W community, the project meetings and workshops were often really motivating. The positive responses to several requests of mine were also very encouraging, such as enabling a ML-Workshop on Zugspitze or just getting Pizza for ECS-seminars.

Last but not least, I am incredibly thankful for my partner Julian. For welcoming me with awesome cocktails after an exhausting workday, for moving to Munich, for providing me with comforting and joyful moments every day - and so much more. Thank you!